THE PROCEEDINGS OF
THE PHYSICAL SOCIETY

VOL. 51, PART 4 1 *July* 1939 No. 286

SPREADING OF PARAFFIN OILS CONTAINING FATTY ACIDS ON AQUEOUS SOLUTIONS OF CALCIUM IONS

By E. H. MERCER

Communicated by Prof. Kerr Grant 21 *November* 1938.
Read in title 24 *February* 1939

ABSTRACT. The spreading of drops of paraffin oil containing stearic acid on alkaline solutions of calcium chloride, during the course of which the acid molecules escape from the oil lens, is described.

§ 1. INTRODUCTION

A PURE paraffin oil, which by reason of the saturated nature of its hydrocarbon constituents does not spread on water, can be made to do so by the addition of small amounts of fatty acids, for instance stearic acid. The molecules of the acid, which diffuse into the oil-water interface, adhere there by means of their active ends, the carboxyl groups, and the long paraffin chains remaining in the oil serve in effect to anchor the oil lens to the water. The phenomena which result when small amounts of such oils are placed on water have been investigated by Langmuir and Blodgett(1, 3). The velocity of spread and the maximum area of the oil lens are influenced greatly by the alkalinity of the underlying water and by the presence of small traces of certain ions, for instance calcium and barium.

Langmuir and Blodgett have determined that on alkaline solutions of pure caustic soda, the strength of which is great enough to ensure that all the fatty acid is finally fixed in the form of a monomolecular layer at the oil-water interface, the law

$$\frac{dA}{dt} \propto A,$$

or

$$\log d \propto t,$$

where A is the area and d the diameter of the oil lens, is closely followed for the greater part of the spread. In the case of paraffin oil containing stearic acid we have confirmed this result, using a cine-camera for rapid spreads. The maximum area covered by the oil, when the alkalinity is sufficient, amounting to 0·01 N., is determined by the number of molecules of fatty acid available, and spreading continues until every molecule has reached the oil-water interface. From this maximum area and the known number of molecules present, the area occupied per molecule in the monolayer can be obtained. It has always been assumed in these calculations that none of the stearic acid has escaped from under the oil on to the surrounding water-surface. The experiments presently to be described show that this is not always true. At the interface between petrolatum and 0·02 N. sodium hydroxide, in the absence of other metallic ions, Langmuir and Blodgett(1, 3) found that the area for the stearic acid molecule is 90 A.^{2}, which may be compared with 20·4 A.^{2}, the area per molecule found in the familiar condensed films of fatty acids on water. Langmuir regards the film between oil and water as gaseous.

The same investigators also found that, when the underlying aqueous solution contained small traces of calcium ions, the area reached by the oil drop was reduced to approximately one quarter of that attained on solutions containing only sodium ions. The area occupied per molecule in the monomolecular film under these conditions then approximates to 20 A.^{2} The film is condensed and apparently solid. An excessively small amount of calcium (sufficient to convert the stearic acid film into a calcium stearate film) brings about this change, and for this reason it is extraordinarily difficult to eliminate entirely the effect of this element when glass vessels are used. Attention(6, 7) has been drawn to the same difficulty in the case of films of fatty acids on a free water surface. The logarithmic law is again followed on 0·01 N. alkaline solutions of calcium salts.

§ 2. FURTHER RESULTS OBTAINED IN THE PRESENCE OF CALCIUM IONS

The thin lenses produced on alkaline water by solutions of stearic acid in paraffin oil are unstable and tend to break up into smaller drops. A drop of such oil placed on water containing only sodium ions first expands up to its maximum area, then breaks up, and finally contracts to a few small drops at the centre, leaving its stearic acid behind on the water surface as a monomolecular film. I have observed more complicated phenomena on solutions containing calcium or barium. When these ions are present it can be shown that the drop begins to shed stearic acid from its periphery while the actual spreading is taking place. There is then beyond the edge of the expanding oil lens a film which spreads ahead of the oil and reaches a final area greater than that of the oil-drop itself. After reaching its maximum area the lens begins to contract as before, leaving the film of stearic acid behind on the water with unchanged area. This resembles the spreading of oleic acid, studied by McBain, Ford and Wilson(4), who showed that the edge of a spreading drop of this substance was preceded by a monomolecular film which rapidly covered the dish and was thereafter compressed by the actual drop itself.

The presence of the monolayer of stearic acid outside the oil drop was shown by photographs obtained in the following way. The oil used was a medium white pure paraffin oil, S1024 (supplied by Vacuum Oil Proprietary Ltd.), to which stearic acid was added in quantities giving concentrations varying from $c = 0{\cdot}0002$ to $c = 0{\cdot}0008$, c being the concentration in grams of acid per gram of oil. Small drops of this solution were placed on weak caustic soda solutions, 0·01 to 0·001 N., which were also 0·001 N. with respect to calcium ions from calcium chloride. Fine aluminium powder was then blown over the surface, and the spreading was photographed with a cinecamera. The photographs show the central oil drop and, surrounding it at a short distance, the edge of the aluminium powder as it is driven back by the expanding film. From the succession of photographs, spaced at definite time intervals, the spreading process can be followed in detail, and graphs can be plotted of area against time.

As can be seen from figure 1, the process is not simple. A drop of oil in which $c=0{\cdot}00084$ when placed on a 0·001 N. caustic soda solution, the concentration of calcium being 0·001 N., flattens under its own weight to a small lens which expands very slowly. At the end of a few seconds the surrounding aluminium powder suddenly moves away from the edge of the lens (indicating the emergence of the stearic acid film) and continues to expand rapidly. The edge of the oil lens follows at a slower rate. The alkalinity used here (0·001 N.) has been chosen for the reason that the slower spread on a solution of dilute alkali enables the details to be seen more clearly, but similar results are obtained on more strongly alkaline solutions; see figure 2. The law

$$\log d \propto t$$

is found to hold both for the oil drop itself and for the edge of the monolayer in the case of more rapid spreads, as in figure 2. For a slow spread, figure 1, the points of both curves can be roughly fitted to the same law over the central part of the spread. If no acid escapes from the lens, then a rate of increase in area proportional to the area is to be expected; but when an appreciable amount of acid reaching the oil-water interface escapes on to the outside air-water surface, it is not clear why the law should still hold, even approximately.

As the alkalinity is decreased from 0·01 N. until the solution is definitely acid, the final area and the average velocity of spread are constantly decreased. This is due to a shortening of the life of a molecule in the oil-water interface, or in other words to a solubility of the adsorbed film in the oil under these circumstances. Drops of strong oil solutions, which on 0·01 N. sodium hydroxide would rapidly spread to an area of many square centimetres, slowly cover a smaller area on solutions of concentrations less than 0·001 N., but steadily shed a large proportion of their acid content on to the water surface; see figures 1 and 2. The irregularities observed on very dilute solutions are probably due to the absorption of carbon dioxide from the air.

On reaching its maximum area the drop breaks up in a manner depending on the rapidity of the preceding expansion. Very slow spreads are followed by a slow contraction, the circular shape being unimpaired; drops which spread in less than five seconds break up in a more complicated fashion. The thin lens breaks into a network of threads, which then collect into a pattern of small lenses. Wilson Taylor has described similar break-up phenomena in some detail(5). The final product in all cases consists of one or several drops of oil resting centrally on a circular area of monomolecular film whose edges are defined by the aluminium powder. If the alkalinity has been sufficiently strong, all the acid is removed from the oil; this is shown by the fact that drops of this oil removed from the surface with a pipette no longer possess the property of spreading.

Experiments have been made to determine the influence that concentration of stearic acid, alkalinity, and temperature have on the amount of acid shed from the expanding oil drop. Figure 2 at *a* and *b* shows the curves obtained by the use of drops of equal size and concentration on 0·01 N. caustic soda, the concentration of Ca^{++} being 0·001 N. at temperatures 17 and 26° C. respectively. The rate of spread increases rapidly with temperature, but the amount of acid that has escaped is much the same in both cases. The effect of varying the alkalinity is still being investigated. As was stated above, the total area of spread is smaller on weakly alkaline solutions, on account of the solubility of the film in the oil at low alkalinities. The amount of acid lost (see figure 1) is relatively greater on such solutions.

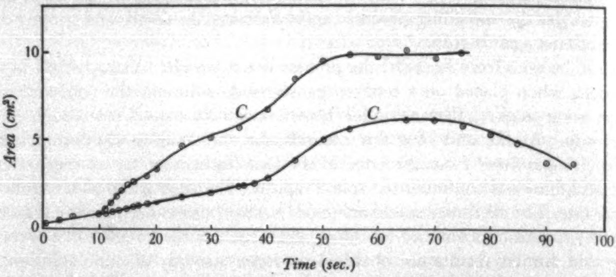


Figure 1. The upper curve corresponds to the edge of the outer film of stearic acid and the lower to the edge of the oil drop. Concentration *c* by weight of stearic acid in oil, 0·00084; concentration of sodium hydroxide, 0·001 N.; of calcium chloride, 0·001 N.; temperature, 17° C.

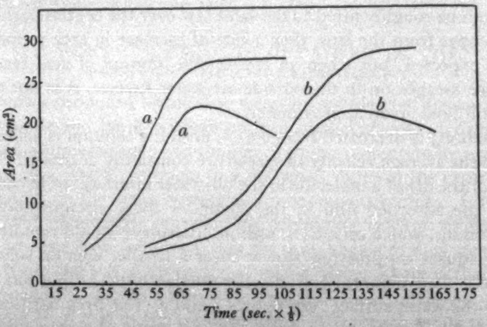
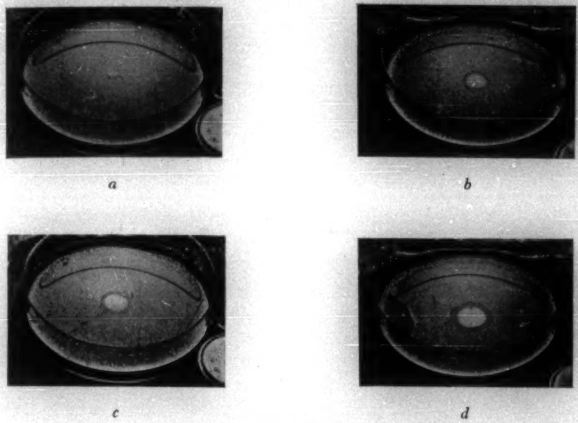


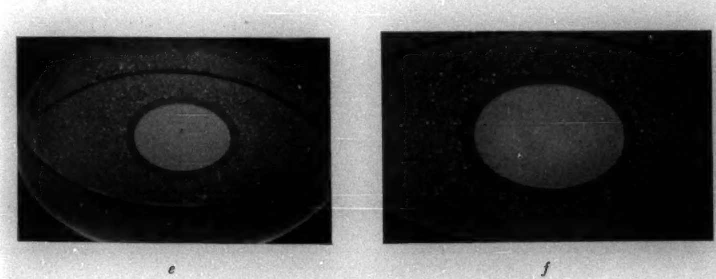
Figure 2. The upper curve corresponds to the position of the edge of the outer film of stearic acid, the lower to the edge of the oil drop. Concentration *c* by weight of stearic acid in oil, 0·00025; concentration of sodium hydroxide, 0·01 N.; of calcium chloride, 0·001 N.; temperature, (*a*) 26° C., (*b*) 17° C.



a b

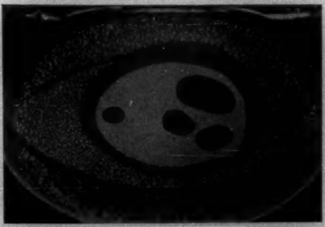
c d

a, *b*, *c* and *d*: successive stages in the spreading of the drop plotted at *C* in figure 1.



e f

e and *f*: enlargements to show details of technique.



g

g: break-up phenomena following a rapid spread.

Spreading on a surface already partially covered by a film of stearic acid deposited from drops of benzene solution is slower, and the surface film which is shed by the expanding oil lens is often visible as a whitish scum consisting apparently of minute drops of oil fastened to the monolayer. On surfaces almost covered by a previous film this is most marked. The oil drop itself does not expand, but from its under edge it feeds a film, covered with globules, into cracks in the solid surface film.

§ 3. MECHANISM OF SPREADING

The mechanism of the spreading process appears to be the following. The molecules of fatty acid diffusing through the oil reach the interface, where their carboxyl groups adhere to the water molecules; the water molecules are in constant heat motion, and diffuse along the surface carrying outwards the adhering fatty acid; the bulk of the oil follows by reason of its adhesion to the hydrocarbon chains of the molecules of fatty acid. This adhesion is very slight, and at the edge of the lens is insufficient to maintain a grip on the expanding surface layer of acid molecules. In the case of slower spreads a close examination of the edge shows that even when the drop as a whole is expanding, the edge itself is losing its hold on the foundation layer and is contracting backward in small puckers. There thus arises a distinct separation between the oil lens and the underlying monolayer. When the outward motion of the surface layer ceases, the oil slowly contracts into a small drop at the centre, often leaving behind small patches of minute oil drops to form a visible scum.

It is possible that in all cases in which the spreading of a neutral oil is due to the addition of an active substance, such as fatty acid, the mechanism is of this nature, although the amount of acid to escape the oil lens during the course of the spreading may be very small. In the case of oils upon 0·01 N. sodium hydroxide containing only sodium or potassium ions, no measurable error is introduced by identifying the area of the oil drop with the area covered by the film of stearic acid; when calcium is present the error may be serious. In the case shown by the figure 1 at *C*, the area of the layer that has escaped the drop at the end of the expansion is nearly twice that of the oil itself.

§ 4. ACKNOWLEDGEMENTS

My thanks are due to the Vacuum Oil Proprietary Ltd. from whom the oils were obtained. I am indebted to the Council for Scientific and Industrial Research for a grant in aid of this research, and to Professor Kerr Grant and Dr R. S. Burdon for suggestions and discussions.

REFERENCES

(1) Langmuir. *J. Franklin Inst.* **218**, 161 (1934).
(2) Langmuir and Blodgett. *Kolloidzschr.* **73**, 257 (1935).
(3) Langmuir. *Science*, **84**, 379 (1936).
(4) McBain, Ford and Wilson. *Kolloidzschr.* **70**, 1 (1937).
(5) Taylor, Wilson. *A New View of Surface Forces* (University of Toronto Press, 1925).
(6) Myers and Harkins. *Nature*, **139**, 367 (1937).
(7) Mitchell, Rideal and Schulman. *Nature*, **139**, 625 (1937).

ULTRA-HIGH-FREQUENCY RESONANCES IN THE POSITIVE-GRID TRIODE

By R. A. CHIPMAN, Acadia University, Wolfville, Nova Scotia

Communicated by J. A. Ratcliffe 16 *November* 1938. *Read in title* 28 *April* 1939

ABSTRACT. A triode with a cylindrical electrode system had its grid maintained at a positive potential, and its plate at a small negative mean potential, with respect to the filament. If an oscillating potential was superposed on the mean plate potential it was found that plate current flowed, and the magnitude of this plate current was determined as a function of the grid potential. There were maxima of plate current for certain values of grid potential, and it was shown that these could be interpreted in terms of a fundamental or harmonic resonance between the period of the applied oscillating potential and the time of transit of the electrons between different pairs of electrodes in the triode.

§ 1. INTRODUCTION

It is commonly observed that electron oscillations generated in a positive-grid triode are accompanied by a flow of plate current. This plate current pulsates at the frequency of the oscillations and is apparently due to the effect on the electron motion in the valve of the generated high-frequency potentials. Gill and Donaldson(1), by applying between the cathode and plate of a positive-grid triode a high-frequency potential from an independent source, have found, correspondingly, a resonant rise in plate current on varying the steady grid potential of the valve. The maximum plate current occurred at a value of grid potential for which the return transit time of electrons between cathode and plate in the valve was approximately equal to the period of the applied high-frequency potential.

This relationship corresponds to the case of fundamental electron oscillations in a valve. Many experimenters, however(2), have obtained dwarf waves of harmonic frequency from electron oscillators, and the author(3) has obtained, from a plane-electrode valve, electron oscillations of period equal to the return transit time of electrons across the grid-plate space alone, and across the cathode-grid space alone. These last two cases seem never to have been observed for a cylindrical triode.

The results described in the present paper have been obtained by using the method of Gill and Donaldson on two types of cylindrical triodes, and show plate-current resonances with the period of the applied high-frequency potential related fundamentally and harmonically to both cathode-plate and grid-plate electron transit times. No clear case of relation to a cathode-grid transit time has been established, and there is now reason to believe from further experiments with the plane-electrode triode that these are not to be expected.

§ 2. EXPERIMENTAL PROCEDURE

The circuit of figure 1 used in the resonance tests is identical with a circuit used by Potapenko(2) in his investigations of the electron-oscillation characteristics of various cylindrical-electrode triodes. With this circuit the high-frequency potential is taken from the oscillator and applied to the triode under test in a perfectly symmetrical arrangement. By varying the length of the grid and plate Lecher-wire systems, it is possible to vary the relative components of high-frequency potential between the different electrode pairs. The high frequency potential was applied to the grid and plate of the test triode because electron oscillators have been found more satisfactory with the external circuit on this pair of electrodes than between cathode and plate.

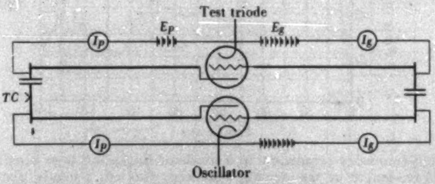


Figure 1. The experimental circuit used for resonance measurements. *TC* is a thermocouple.

The oscillator valve used throughout was a triode of AT 40 type operated as an electron oscillator at wave-lengths of about 80 cm. The experimental procedure, as with Gill and Donaldson, was to keep the high-frequency potential applied to the test triode constant, vary the steady grid potential from zero to several hundred volts, and observe the variations of plate current. The plate of the test triode was kept always at zero or a low negative potential with respect to the negative end of the filament. The principal precautions necessary were concerned with keeping the high-frequency potential constant in magnitude and frequency, and keeping the filament temperature constant during any one experiment. These required running the oscillator for some time before each series of observations, and the use of steady sources for grid and filament potentials. A constant check was kept on oscillator frequency. The steadiness of the applied high-frequency potential was assumed to be related to the constancy of the current passing a thermocouple in the bridge of the plate Lecher-wire system, and to the constancy of the oscillator plate current. With no high-frequency potential applied to the electrodes, the plate current of the test triode was in all cases negligibly small over the whole range of grid potential.

§ 3. EXPERIMENTAL RESONANCE CURVES

Figures 2, 3 and 4 show some of the plate-current curves obtained. These are typical curves and illustrate most of the phenomena involved.

For figure 2 the test triode was an *R* valve. With the grid current temperature-limited to a steady value of 2·8 ma., space-charge saturation existed only for grid potentials up to about 10 v. A negative plate potential of 2·0 v. relative to the negative end of the filament was used to stop the small plate current at zero high-frequency potential.

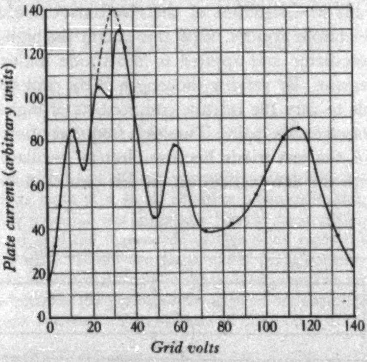


Figure 2. Ultra-high-frequency resonances in a triode of English *R* type obtained with the circuit of figure 1. Wave-length of the driving potential, 76·5 cm.; steady plate voltage, −2·0 v.; filament potential constant, giving a temperature-limited emission of 2·8 ma., and providing space-charge saturation for grid potentials up to 10 v.

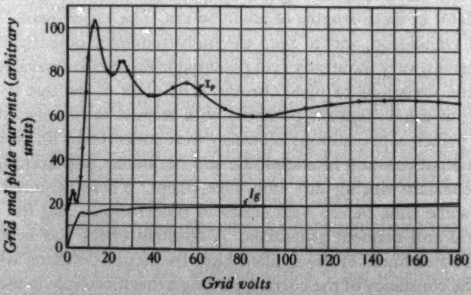


Figure 3. Ultra-high-frequency resonances in a triode of AT 40 type. Wave-length of driving potential, 75·8 cm.; plate potential, zero; temperature-limited filament-emission, 0·2 ma.

The test triode for figure 3 was of AT 40 type, of larger dimensions than the *R* valve. The temperature-limited grid current was 0·20 ma. providing space-charge saturation for grid potentials up to about 5 v. The plate potential was zero. Figure 4 is for the same test triode as figure 3, showing the effect of three different filament-temperatures, the highest providing space-charge-limited grid current up to about 50 v. grid potential. At higher filament-temperatures the test triode broke into oscillation. The plate potential was again zero.

It was found impossible to make any reasonable estimate of the relative values of the high-frequency potentials between the different pairs of electrodes. It was, however, certain that these could be varied by varying the two Lecher-wire systems, and this was done. Figures 2, 3 and 4 were taken for three arbitrarily different distributions of high-frequency potential. The maximum value of the high-frequency potential generated by the oscillator was measured with some accuracy by increasing the negative plate potential on the test triode until no plate current was observed for any adjustment of the Lecher-wire systems. This showed a peak value of about 4 V.

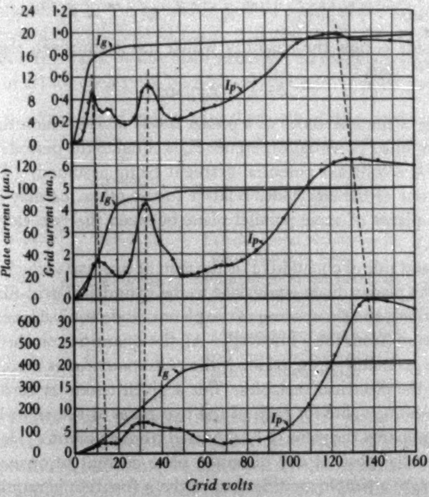


Figure 4. Ultra-high-frequency resonances in a triode of AT40 type, showing the effect of increasing space-charge density. Wave-length of the driving potential, 80·4 cm.; plate potential, zero. The dotted lines represent an attempt to trace the resonance peaks from one curve to the next.

§4. BASIS OF ANALYSIS OF EXPERIMENTAL RESONANCE CURVES

For the analysis of the curves of figures 2, 3 and 4 in terms of transit-time phenomena, the period of the applied high-frequency potentials is known accurately. The problem is thus one of determining the transit time corresponding to the grid voltage at each resonance peak and comparing this with the known period.

Scheibe(4) has derived an equation for the wave-length of electron oscillations generated in a cylindrical triode in the absence of space charge as follows:

$$\lambda=\frac{4cr_1}{\sqrt{(2\,e/m\,E_g)}}\left[f\left\{\sqrt{\left(\log_e\frac{r_1}{r_0}\right)}\right\}+g\left\{\sqrt{\left(\frac{E_g}{E_g-E_a}\log_e\frac{r_2}{r_1}\right)}\right\}\right],\quad\text{......(1)}$$

where r_0, r_1, r_2 are the radii of filament wire, grid cylinder, and plate cylinder respectively, E_g, E_a are the steady grid and plate potentials, c is the velocity of light, and e/m is the specific charge of the electron. The functions f and g are conveniently tabulated by Kapzov and Gwosdower(5) for the range of the variables likely to be encountered experimentally. This equation assumes that the period of the oscillations is equal to the transit time of an electron from the filament to the plate and back.

Equation (1) can be rewritten as

$$\frac{\lambda\sqrt{(Eg)}}{2023}=r_1\left[f\left\{\sqrt{\left(\log_e\frac{r_1}{r_0}\right)}\right\}+g\left\{\sqrt{\left(\frac{E_g}{E_g-E_a}\log_e\frac{r_2}{r_1}\right)}\right\}\right]. \qquad \text{......(2)}$$

The left-hand side is now the quantity which has been given the symbol δ in a previous paper(3); thus

$$\delta=\lambda\sqrt{(Eg)}/2023. \qquad \text{......(3)}$$

It is the spacing of the electrodes of a planar diode across which the return transit time of electrons under potential-difference E_g is equal to the period of the oscillation of wave-length λ, a constant potential-gradient being assumed. (The potential E_g may be that of one of the electrodes relative to the other, or may be the potential relative to both planes of some parallel plane between the two electrodes, as in an idealized triode.)

The right-hand side of equation (2) is the sum of two parts which are respectively the spacing of the planar diode equivalent to the cylindrical grid-filament structure (from the f function), and the spacing of the planar diode equivalent to the cylindrical grid-plate structure (from the g function). All the quantities on the right-hand side involve only the geometry of the triode if the plate potential is zero, while those on the left involve the electrical variables. For a given triode at zero plate potential there will be a unique value for each of the two parts of the right-hand side, with which may be compared the value of δ calculated from equation (3) for any generated oscillation or for the peak of any observed plate-current resonance. If the plate potential is not zero a simple modification of the g function is required.

§ 5. ANALYSIS OF EXPERIMENTAL CURVES

Table 1 gives the dimensions of the two triodes used in the resonance experiments, and shows the values of the constants calculated from the Scheibe equation for these dimensions, zero plate potential being assumed. The letters A, B and C refer respectively to the grid-plate constant, the filament-plate constant, and the filament-grid constant.

Table 1. Dimensions and Scheibe constants of valves (cm.)

	r_0	r_1	r_2	$r_1 f$ or C	$r_1 g$ or A	B or $A+C$
R valve	$0\cdot004_0$	$0\cdot22_5$	$0\cdot50_0$	$0\cdot13_5$	$0\cdot32_6$	$0\cdot46_1$
AT 40	$0\cdot008_0$	$0\cdot24_0$	$0\cdot64_0$	$0\cdot14_9$	$0\cdot47_2$	$0\cdot62_1$

The grids of both triodes were helices of very uniform diameter, and both grids were very closely coaxial with the plates. The plate cylinders were fairly uniform, but were closed only by an over-lap, so that the value of r_2 varied slightly from point to point. The probable error in A is not greater than about 3 per cent. The filaments in both cases were not appreciably eccentric as a whole but were not perfectly straight. In the AT 40 valve, particularly, the filament strayed by a considerable fraction of a millimetre at some places from the axis of the electrodes. This means a moderately large but indeterminate possible error in C, which will lead to a small error in B. C is probably accurate within 10 per cent for the R valve and 20 per cent for the AT 40 valve, both of the figures in table 1 being upper limits.

Table 2 represents the analysis of figure 2 by the method described in the preceding section. For this table it has been necessary to make a separate calculation of A (and hence of B) for each peak because of the negative plate potential.

Table 2. Analysis of figure 2 by the method discussed in § 4

(1) E_g at peak	(2) δ	(3) Suggested comparison	(4) Meaning of comparison	(5) Deviation of (3) from (2) (per cent)
$115{\cdot}5 \pm 1{\cdot}0$	$0{\cdot}40_7$	$B = 0{\cdot}44_1$	Fundamental resonance in plate-filament space	8
$57{\cdot}5 \pm 0{\cdot}5$	$0{\cdot}28_6$	$A = 0{\cdot}29_8$	Fundamental resonance in grid-plate space	4
$30{\cdot}5 \pm 0{\cdot}5$	$0{\cdot}20_9$	$B/2 = 0{\cdot}21_1$	1st harmonic resonance in plate-filament space	1
$11{\cdot}7 \pm 0{\cdot}3$	$0{\cdot}12_7$	$B/3 = 0{\cdot}12_8$	2nd harmonic resonance in plate-filament space	1

The analysis shows that the first, second and fifth peaks from the right in figure 2 have values of δ corresponding with excellent accuracy to B and its first two submultiples respectively, indicating fundamental and harmonic resonances between the period of the applied high-frequency potential and the filament-plate electron transit time. The third and fourth peaks from the right do not directly fit the resonance analysis at all well, but if they are regarded as being the two sides of a single split peak the result is satisfactory, and δ for this peak corresponds closely to A. This effect was encountered in the resonance experiments on other occasions, and may be due to some instability that arises with too high a plate-current flow.

Table 3 gives a corresponding analysis of the resonance peaks of figure 3. Its interpretation is similar to that for table 2 with two exceptions. As in table 2 there are harmonic resonances with the filament-plate transit time, being in this case the first, second, and third harmonic orders. But the fundamental of this series, which should appear at a grid potential of 220 v., is absent. (This value of grid potential does not appear on the curve, but all curves were carried up to 350 v.) A theory developed by Alfven⁽⁶⁾ has suggested that under certain conditions harmonic oscillations might be generated across the filament-plate space of a triode and the fundamental of these be unobtainable.

Table 3. Analysis of figure 3 by the method discussed in § 4

(1) E_g at peak	(2) δ	(3) Suggested comparison	(4) Meaning of comparison	(5) Deviation of (3) from (2) (per cent)
150 ± 10	0·45	$A = 0{\cdot}46$	Fundamental resonance in grid-plate space	2
54 ± 2	0·28	$B/2 = 0{\cdot}30$	1st harmonic resonance in filament-plate space	7
25·0 ± 0·5	$0{\cdot}18_9$	$B/3 = 0{\cdot}20_3$	2nd harmonic resonance in filament-plate space	7
12·9 ± 0·1	$0{\cdot}13_5$	$B/4 = 0{\cdot}15_2$	3rd harmonic resonance in filament-plate space	12
2·4 ± 0·1	$0{\cdot}05_7$	?	?	—

The fundamental resonance with the grid-plate transit time is again present, and the deviation for this resonance is small as before. There is no harmonic resonance with the grid-plate transit time unless the peak at 12·9 v. should be considered as belonging to this category rather than to the filament-plate series. The coincidence of two harmonics here might account for the very high value of plate current at the 12·9 v. peak.

The resonance peak at grid potential of 2·4 v. occurs in a region of complete space-charge saturation. Its significance remains entirely obscure. It does not seem satisfactory to regard it as a harmonic resonance of high order when the intermediate orders are not present.

The correlation of columns (2) and (3) of table 3 is for the last four peaks as satisfactory as in table 2 only if the value of C is taken as about 20 per cent less than that calculated from the formula, which is explicable on the basis of the somewhat rambling filament of this particular triode.

The upper curve of figure 4, taken at low filament temperature, is effectively free from space-charge effects, and can be analysed similarly to figures 2 and 3. This analysis is given in table 4.

Table 4. Analysis of the upper curve of figure 4 by the method of § 4

(1) E_g at peak	(2) δ	(3) Suggested comparison	(4) Meaning of comparison	(5) Deviation of (3) from (2) (per cent)
124 ± 1	$0{\cdot}44_5$	$A = 0{\cdot}47_2$	Fundamental resonance in grid-plate space	6
35·0 ± 0·5	$0{\cdot}23_5$	$A/2 = 0{\cdot}23_6$	1st harmonic resonance in grid-plate space	<1
15·9 ± 0·3	$0{\cdot}15_7$	$A/3 = 0{\cdot}15_7$	2nd harmonic resonance in grid-plate space	<1
8·7 ± 0·2	$0{\cdot}11_8$	$A/4 = 0{\cdot}11_8$	3rd harmonic resonance in grid-plate space	<1

The triode used here was the same as that used for figure 3, and a similar analysis might be expected. However, the lengths of the grid and plate Lecher-wire systems were different in the two cases, and this seems to have changed the location of the resonance mechanism. Table 4 shows no resonances with a filament-plate transit time, but a fundamental and three harmonic resonances with the grid-plate transit time.

Considering the three parts of figure 4 together, it is obvious that the effect of space charge on transit time is not a simple one. For the peak at highest grid potential, the increasing space charge has tended to increase the transit time. For the peak around 35 v. the effect is a slight decrease of transit time. The effect on the other peaks is very confused. The results cannot be said to offer any confirmation of simple theory as presented, for example, by Cockburn(7). The fact that the peaks tend to be smoothed out in the region of complete space-charge saturation may be consistent with a previous conclusion(3) that the mechanism of space-charge oscillations is essentially different from that of negative-resistance oscillations.

§ 6. CONCLUSION

It is to be noted that the transit-time calculations with which the oscillation periods have in effect been compared, through the equivalent planar diode analysis, take into consideration only the cylindrical shape of the electrodes, and ignore a number of other factors which might be expected to influence transit time. These include (*a*) space charge; (*b*) the fall of potential between grid wires; (*c*) deflection of electrons at the grid; (*d*) potential drop along the filament; (*e*) initial electron velocity; (*f*) high-frequency potentials on the electrodes.

Consideration suggests, however, that those electrons which reach the plate will be the ones least affected by factors (*c*) and (*d*), while the effects of factors (*a*), (*b*) and (*e*) may be expected to be very small under the conditions of the resonance experiments. Independent analysis has indicated the effect of (*f*) to be greatly reduced as the harmonic order of resonance increases. Where the harmonic order is low the steady grid potential is high, and a high-frequency potential not greater than 10 per cent of this steady potential will not have an appreciable effect. The very close agreements in table 4 may therefore be taken as significant of close relationship between period of oscillation and transit time.

If these plate-current resonances are a result of the same mechanism that underlies electron oscillations, they imply that the valves used in the tests might be capable of generating dwarf waves of several orders, and oscillations having periods related to a grid-plate electron transit time. The fact that numerous attempts failed to produce any evidence of these may be ascribed to low efficiency of the mechanism under such conditions and to the increase at higher frequencies of losses in the external circuit and of radiation losses.

The apparent relative enhancement of the 8·7-v. peak in the upper curve of figure 4 and the 35-v. peak in the centre curve is relevant, because of the peaks being near the upper knee of the grid-current saturation curves, to the experimental fact that cylindrical triodes invariably oscillate most efficiently when adjusted to this point on their grid-current characteristic.

§ 7. ACKNOWLEDGEMENTS

The experimental work described in this paper was carried out in the Cavendish Laboratory of Cambridge University. I wish to express my thanks to Mr J. A. Ratcliffe for advice and inspiration and for much helpful criticism in the preparation of the paper.

REFERENCES

(1) Gill, E. W. B. and Donaldson, R. H. *Phil. Mag.* **15**, 1177 (1933).
(2) Potapenko, G. A. *Phys. Rev.* **39**, 625 and 638 (1932).
(3) Chipman, R. A. *Proc. Phys. Soc.* **47**, 1042 (1935).
(4) Scheibe, A. *Ann. Phys., Lpz.*, **73**, 54 (1924).
(5) Kapzov, N. and Gwosdower, S. *Z. Phys.* **45**, 114 (1927).
(6) Alfven, H. Uppsala University Dissertation (1934).
(7) Cockburn, R. *Proc. Phys. Soc.* **47**, 810 (1935).

ON THE THEORY OF ELECTRONIC SEMI-CONDUCTORS

BY B. R. A. NIJBOER, H. H. Wills Physical Laboratory,
University of Bristol

Communicated by Prof. N. F. Mott 29 *June* 1938 *and in amended form* 23 *January* 1939.
Read in title 10 *March* 1939

ABSTRACT. A discussion of the discrepancies between Wilson's theory of semi-conductivity and recent experiments on the subject is given; and on the basis of Wagner and Schottky's theory of *Fehlordnung* an extension of Wilson's theory, which may remove the discrepancies in some cases, is proposed.

§ 1. INTRODUCTION

ACCORDING to Gudden(1) and Wagner(2) we must assume that in most semi-conductors the conductivity is caused entirely by deviations from stoichiometric composition. In some substances, such as Cu_2O, NiO, UO_2 and the cuprous-halides, there is an excess of oxygen (or halogen) in the crystal and the conductivity σ increases with increasing pressure of the electronegative constituent; in others, like ZnO, CdO, Al_2O_3 and Ta_2O_5, σ decreases, and it is probable that in the crystal there is an excess of metal.

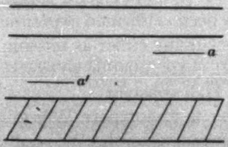


Figure 1. Energy levels for a semi-conductor.

Wilson(3) has given a quantum-mechanical theory of semi-conductivity. He considers those substances that have a fully occupied energy band and an empty one separated by a gap, which will usually be too big to account for an appreciable current. Between the two bands impurity levels *a*, figure 1, may be situated, which supply the upper band with free electrons by thermal activation. It is also possible that levels *a′* may receive electrons from the full band. The positive holes in the lower band will carry the current in the latter case. Here we should expect an anomalous (positive) sign of the Hall constant *R*, as has been observed, for instance, for Cu_2O(4) and UO_2(5).

The theoretical formula for the Hall coefficient *R* is

$$R = \pm 3\pi e/8nc,$$

where n is the number of free electrons (or holes) per unit volume. Thus the Hall coefficient gives a direct measure of the number of free electrons.

Wilson obtains the following formula for the number of electrons in the conduction band

$$n=\left(\frac{n_0}{4\pi^3}\right)^{\frac{1}{2}}\left(\frac{2\pi mkT}{\hbar^2}\right)^{\frac{3}{4}} e^{-\Delta E/2kT}. \qquad \text{......(1)}$$

Since the conductivity σ of a semi-conductor may be written thus:

$$\sigma=nv, \qquad \text{......(2)}$$

where v is the mobility of an electron, and since we should not expect any rapid variation of v with T, it follows that we may write

$$\sigma=Ae^{-\epsilon/kT}, \qquad \text{......(3)}$$

where A varies only slowly with T. In many cases the experimental values of σ can be expressed by a formula of this type; we then, according to Wilson's theory, have to identify ϵ with $\frac{1}{2}\Delta E$. One purpose of this paper is to discuss this factor $\frac{1}{2}$, and to show that cases exist in which it is correct to identify ΔE with ϵ.

In connexion with their theory of *Fehlordnung* Wagner and Schottky⁽⁶⁾ have developed the following picture of the nature of the impurity levels. Consider, for instance, an "oxidation semi-conductor" such as Cu_2O whose conductivity increases with excess of oxygen. The crystal carries its excess of oxygen by having Cu lattice points whence the Cu^+ ion is missing, the $O^=$ ion being too big to occupy an interlattice position. As the crystal must be electrically neutral, electrons must be removed too. The electron will be removed either from the Cu^+ ion or from the $O^=$ ion next to the hole. This will give rise to electron levels of the type a' above the full Cu^+ or $O^=$ band. These oxidation semi-conductors will therefore show a positive Hall constant, as has been confirmed experimentally for Cu_2O and UO_2. If there is an excess of metal, occurring either as missing oxygen ions or as metal ions on interstitial positions (ZnO), there must be an excess of electrons too, giving levels of the type a and a negative Hall constant.

One purpose of this paper is to compare Wilson's theory with experiment, in view of this model for the impurity levels. We shall briefly discuss (in § 2) some experiments made on the subject by different authors in the last few years, and shall see that many of the results obtained cannot be understood in terms of Wilson's theory without some modifications. Some of the anomalies may be explained by a consideration of the factor $\frac{1}{2}$ in the activation energy as already mentioned. But, on the other hand, the very large variations with concentration of the activation constant ϵ for TiO_2, Fe_2O_3 and ZnO recently found by Meyer and Neldel⁽⁷⁾, will probably need quite new conceptions for their explanation.

§ 2. THE EXPERIMENTAL DATA

Engelhard⁽⁴⁾ has made very elaborate experiments on the conductivity and Hall constant of Cu_2O for temperatures between −130 and +130° c. The conductivity is found to be due to the excess of oxygen in the crystal. The Hall constant is positive, thus indicating conductivity due to holes. Engelhard plotted log R and the logarithm of the resistance as functions of $1/T$ for several specimens (see his figures on pp. 520 and 521). His graphs are straight lines with slopes depending only slightly on the previous treatment. They show, however, two kinks. The one at room-temperature (K_1) shows inertia, as had been described already by Vogt⁽⁸⁾. Engelhard and Vogt ascribe it to a phase transition, the more so as the temperature-dependence of the electron mobility (as given by $R\sigma$) also shows the kink. The other one (K_2) at −60° c. is attributed by Engelhard to a continuous range of binding energies, as given by the infra-red absorption, determined by Schönwald⁽⁹⁾ for the same specimen (see also § 3). Engelhard's results seem to indicate on comparison of the infra-red and conductivity measurements that the slopes of the curves for Cu_2O are given by ΔE, instead of $\frac{1}{2}\Delta E$, as Wilson's theory demands.

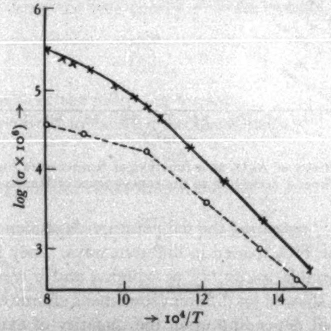


Figure 2. Conductivity of ZnO as a function of temperature: ——, according to von Baumbach and Wagner; ----, according to Jander and Stamm.

Kröger⁽¹⁷⁾ has also made measurements of the temperature dependence of the conductivity of Cu_2O. For untempered specimens he finds the average value 0·21 ev. for the constant ϵ. For specimens which are tempered in oxygen at 950° c. the average value of ϵ is $0{\cdot}10_5$ ev.; for those tempered in vacuum 0·26 ev. Waibel⁽¹⁸⁾ has found results qualitatively similar; the variations of ϵ are bigger, however, amounting to 0·20, 0·06 and 0·23 ev.

Jander and Stamm⁽¹¹⁾ and afterwards von Baumbach and Wagner⁽¹²⁾ have measured the conductivity of ZnO as a function of temperature between 400 and 1000° c. The conductivity decreases with increasing oxygen pressure. The {log σ, temperature} curve shows an anomaly at about 700° c., figure 2. The possibility of a phase transition has been eliminated by Ebert⁽¹³⁾ by means of x-ray measurements.

Hartmann⁽⁵⁾ has investigated the conductivity of Al_2O_3 and Ta_2O_5 as functions of temperature between 20 and 400° c. Pure Al_2O_3 is an insulator. After several hours of heating in vacuum the conductivity becomes measurable. By again plotting log σ as a function of $1/T$, Hartmann finds a straight line with ϵ equal to 0·385 ev.

After further heating more oxygen atoms were removed, the number n_0 of excess metal atoms became bigger, and finally an equilibrium state was reached. He then found another straight line shifted towards higher conductivity, but with ϵ equal to 0·25 ev. By subsequent heating in oxygen he obtained straight lines with conductivities and slopes between the former two values. Reduction with hydrogen again had the effect of increasing the conductivity, figure 3. Qualitatively the same applied to Ta_2O_5.

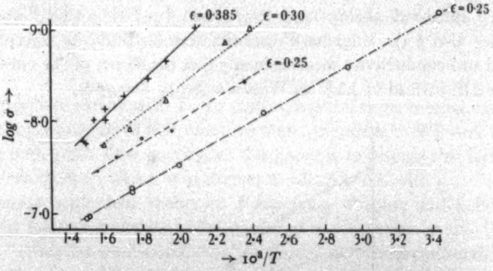


Figure 3. Conductivity of Al_2O_3 as a function of temperature for different treated specimens, according to the experiments of Hartmann.

Meyer and Neldel(7) measured the temperature-dependence of the conductivity for many specimens of TiO_2 treated in different ways. They found that σ could be represented by the formula $\sigma = ae^{-\epsilon/kT}$ in which ϵ and a were constants for each specimen. ϵ, however, showed for different specimens enormous variations (between 1·73 and 0·004 ev.), all depending upon the quantity of excess metal. After the reduction of a given substance with hydrogen, ϵ decreases. Moreover, for values of ϵ which are not too small (i.e. $>0·06$ ev.) they found ϵ to be a linear function of log a. The same applied for ZnO and Fe_2O_3.

Finally we only mention the results of Kapp and Treu(10) who found for CuO values of ϵ varying between 0·12 and 0·24 ev., without any obvious connexion with the conductivity.

§ 3. THEORETICAL DEVELOPMENTS*

In addition to the vacant lattice points due to a stoichiometric excess of either constituent in a polar lattice, there are, according to Wagner and Schottky(6), a number of vacant lattice points of both constituents in any crystal in thermal equilibrium. Their number N will be given by

$$N = N_0 e^{-W/kT}. \qquad \text{......(4)}$$

---

* After this paper had been finished, it came to our notice that considerations similar to those given here in § 3 had been given already (though not on the basis of the *Fehlordnung* theory) by de Boer and van Geel(16). These authors, however, did not compare their results with experiments in any detail.

Here N is the number of ions of a given kind per cm.^{3} and is of the order 10^{22}; and for simple crystals, e.g. for alkali halides, $W=\frac{1}{2}(W_+ + W_- - W_L)$, where W_+ and W_- are respectively the energies requisite to remove a positive and a negative ion out of the crystal, and W_L is the lattice energy per ion pair. As the crystal is cooled a certain number of these holes will be frozen in. These frozen-in holes are exactly the same as the impurity levels due to stoichiometric excess of metal; they are not, however, normally occupied by electrons.

We must thus take for our model of a semi-conductor a crystal with, per unit volume, n_0 excess atoms of one kind, and N frozen-in vacant lattice points. There are thus n_0 electrons (or holes) available and n_0+N possible impurity levels in which an electron may be held. We require for this case a formula for the number n of electrons in the conduction band at any temperature, corresponding to formula (1) for the case in which $N=0$.

We have to write down the free energy F of the system, and determine n from the condition that

$$\left(\frac{\partial F}{\partial n}\right)_T = 0. \qquad \text{......(5)}$$

The free energy consists of the following terms:

(1) The energy required to excite n electrons, namely

$$n\,\Delta E.$$

(2) The free energy of the electron gas, namely

$$-nkT\left[\log\left(\frac{2\pi mkT}{h^2}\right)^{\frac{3}{2}} + \log\frac{V}{n} + 1\right],$$

where V is the volume.

(3) The free energy of $(n+N)$ holes distributed among (n_0+N) possible levels; this gives

$$-kT\log\frac{(N+n_0)!}{(n+N)!\,(n_0-n)!}.$$

Applying formula (5) we obtain for n

$$\frac{n\,(N+n)}{n_0-n} = V\left\{\frac{2\pi mkT}{h^2}\right\}^{\frac{3}{2}} e^{-\Delta E/kT}. \qquad \text{......(6)}$$

This formula is exact. In the case in which N, the number of frozen-in holes, is small compared with the number of available electrons, there is a rather sharp change in the slope of the $\{\log n,\ 1/T\}$ lines. From equation (6) one finds, per unit volume,

$$n = \frac{n_0}{N}\left\{\frac{2\pi mkT}{h^2}\right\}^{\frac{3}{2}} e^{-\Delta E/kT}, \quad T \ll T_c \qquad \text{......(7}a\text{)}$$

$$= n_0^{\frac{1}{2}}\left\{\frac{2\pi mkT}{h^2}\right\}^{\frac{3}{4}} e^{-\frac{1}{2}\Delta E/kT}, \quad T \gg T_c \qquad \text{......(7}b\text{)}$$

where T_c is given by $n=N$, and hence by

$$N = \left(\frac{n_0}{2}\right)^{\frac{1}{2}}\left\{\frac{2\pi mkT_c}{h^2}\right\}^{\frac{3}{4}} e^{-\frac{1}{2}\Delta E/kT_c}.$$

Provided that the critical temperature falls within the measured temperature-range, we expect a result as shown schematically in figure 4.

The anomaly of ZnO discussed in § 2 may be well explained by this mechanism. In the experiments of von Baumbach and Wagner the oxide was formed at 1000° C. As only the electronic conductivity σ, equal to nv where v is the electron mobility, has been measured, we have no direct knowledge of n. Measurements of the Hall constant R are better suited for comparison with the theory.

The Cu_2O examined by Engelhard was formed at 1000° C. The phenomenon of inertia rules out the possibility of ascribing the kink K_1 to the presence of the frozen-in holes. Most probably it must be attributed to a transition, as Vogt and Engelhard assumed. Kröger's and Waibel's results suggest (see below) that Engelhard worked in the region where $\epsilon = \Delta E$, so that we expect the critical temperature of figure 4 to lie above the measured range. Now Wagner and his collaborators(15) have measured the ionic conductivity of Cu_2O at 1000° C., the temperature of formation of Engelhard's specimens. The number of missing Cu^+ ions, which is responsible for the current, is $(N_0 e^{-W/kT} + n_0)$. Let their mobility be $V_0 e^{-A/kT}$. Assuming very roughly that $W = A$ (as is the case for the alkali halides(14)) and taking n_0 equal to 10^{20}, we find from the observed ionic conductivity that

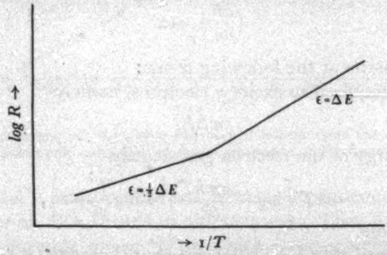


Figure 4. Hall constant as a function of temperature in the case of frozen-in holes.

$$W = 1{\cdot}7 \text{ ev.}; \quad N \fallingdotseq 10^{16}.$$

This would yield a critical temperature T_c approximately equal to 200° C., outside the range measured by Engelhard. So we have in the measured range (form 7(a))

$$n = \frac{n_0}{N}\left\{\frac{2\pi mkT}{h^2}\right\}^{\frac{3}{2}} e^{-\Delta E/kT}.$$

For a given temperature n is known from the measured Hall constant. By taking again n_0 equal to 10^{20} we find from this formula again $N \fallingdotseq 10^{16}$, which thus confirms our model. Thus it seems very probable that the factor $\frac{1}{2}$, as given by Wilson's theory, does not come in for Cu_2O in the range of temperatures measured by Engelhard.

By tempering Cu_2O in oxygen, n_0 the number of excess oxygen ions is made bigger, N remaining constant. The critical temperature shifts towards lower temperatures. If it comes below the low-temperature end of the measured range, we expect to find a straight line with slope $\frac{1}{2}\Delta E$, whereas if it lies higher than the high-temperature end (which it may do after the Cu_2O has been heated *in vacuo*) we shall get a straight line with ϵ equal to ΔE. If $N \fallingdotseq n$ for a temperature within the measured range, say near to the high-temperature end, we shall get a set of experimental points not deviating within the experimental error from a straight line, but the latter will give a slope somewhat smaller than ΔE.

This would be quite satisfactory as an explanation of Kröger's results, § 2. Of course, we cannot expect ϵ to change exactly by a factor 2, because the mobility also depends on temperature; the factors $T^{\frac{3}{4}}$ and $T^{\frac{3}{2}}$ will also have some influence. The experiments of Waibel, on the other hand, yield variations which may be too big to allow the foregoing interpretation to be correct here. We are led to the same conclusion if we consider that the ϵ-values for Cu_2O obtained by different authors spread so considerably.

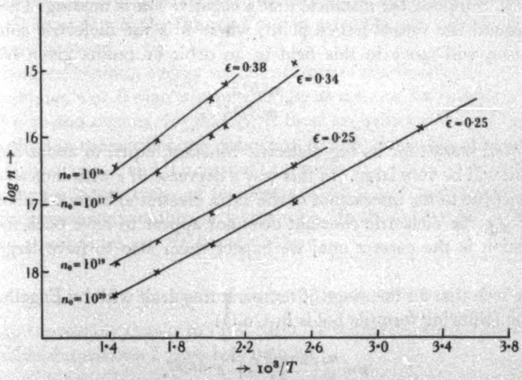


Figure 5 (theoretical). The number of free electrons in Al_2O_3 as a function of temperature for specimens with different amounts of excess metal.

In Hartmann's experiments on Al_2O_3, the number n_0 of excess metal atoms is decreased by heating in oxygen and increased by heating *in vacuo* or in hydrogen. Apart from these facts, the previous observations apply again. In order to show that the mechanism of the frozen-in holes is able to fit his results for Al_2O_3, we calculated n from formulae (7), by taking ΔE equal to 0·45 ev. and assuming reasonable values for N, n_0, namely, $N = 10^{16}$, $n_0 = 10^{16}$, 10^{17}, 10^{19}, 10^{20}. In figure 5 we have plotted $\log n$ as a function of $1/T$, and we find a quite good agreement with Hartmann's experimental results, figure 3. The fact that in the experimental diagram the lines lie closer together may be due to a decrease in the mobility with increasing number of impurities.

The experiments of Meyer and Neldel, § 2, make it difficult, however, to maintain this explanation. It is remarkable that the extremely small values of ϵ have only been found for very big concentrations of impurity in substances in which the conductivity is due to an excess of metal, namely TiO_2, ZnO and Fe_2O_3. This suggests that somehow a formation of colloidal metal in the oxide may be responsible for the observed small values of ϵ. But even if we exclude these very small binding energies, there remain variations of ϵ amounting to a factor of about 5 and, moreover, the linear relation between ϵ and the quantity log $n_0 v$, where v is the mobility, are facts which cannot be understood at all on the basis of our extensions of Wilson's theory. It is difficult to see how an interaction of the impurities for moderate concentrations could account for this behaviour.

A solution of the difficulty might, however, be found on the following lines (cf. Mott(19)). Suppose, for instance, that a negative ion is missing. There will be a field e/kr^2 round the vacant lattice point, where k is the dielectric constant. The excess electron will move in this field in an orbit or radius given by the Bohr formula

$$r=\frac{n^2\hbar^2 k}{mZe^2}.$$

TiO_2 is well known for its big dielectric constant, equal to about 80, so in this case the orbit will be very large. In this way a decrease of ϵ with increasing number of impurities (due to the interaction of the extra electrons) becomes plausible. For ZnO and Fe_2O_3 the dielectric constant does not appear to have been measured; if our explanation is the correct one, we expect them also to have large dielectric constants.

We have seen that for the range of temperatures dealt with by Engelhard(4) most probably the following formula holds in Cu_2O:

$$n=\frac{n_0}{N}\left(\frac{2\pi mkT}{h^2}\right)^{\frac{3}{4}} e^{-\Delta E/kT}.$$

In order to explain the kink k_2 at −60° c. he assumes more than one binding energy. If we suppose that n_{01} electrons per cm.^{3} have binding energies ΔE_1, and n_{02} a binding energy ΔE_2, we may write

$$n=\frac{n_{02}}{N}\left(\frac{2\pi mkT}{h^2}\right)^{\frac{3}{4}} e^{-\Delta E_2/kT}\left(1+\frac{n_{01}}{n_{02}}\,e^{-(\Delta E_1-\Delta E_2)/kT}\right),$$

and it is obvious that we shall have a kink at the temperature for which

$$\frac{n_{01}}{n_{02}}\,e^{-(\Delta E_1-\Delta E_2)/kT}\doteq 1.$$

Engelhard is able to fit his experiments by assuming that

$$n_{01}=10^{20},\quad \Delta E_1=0{\cdot}36 \text{ ev.},\quad n_{02}=5{\cdot}10^{15},\quad \Delta E_2=0{\cdot}18 \text{ ev.}^*$$

---

* The reason why his fourth specimen yields quite different values may be that it shows a transition kink, which causes a deviating value of n at 333° c. (cf. p. 524 of his paper).

He believes, however, that the true representation of the facts must be sought in the continuous range of binding energies such as that given by the infra-red absorption determined by Schönwald⁽⁹⁾. In our view there are serious objections to this view. First, it is difficult to see how it could give rise to a kink and not to a continuously increasing slope towards higher temperatures. Engelhard's underlying calculations cannot be given too much weight, because they depend sensitively on the shape of the long-wave-length tail, in which $\epsilon < 0{\cdot}4$ ev., of the absorption curve, and this cannot be determined very accurately. Moreover, as has been pointed out by de Boer and van Geel⁽¹⁶⁾, the excitation energy for thermally and optically released electrons cannot be expected to be the same. After an electron has been removed, the surrounding ions will go into new positions in consequence of the field of the positive hole left by the electrons, so that polarization energy will be released. The latter should, however, not be taken into account for optically released electrons, because here, according to the Franck-Condon principle, the ions do not move until the absorption is over.

More than one binding energy would, of course, also be able to give an explanation for Hartmann's or Kröger's results.* Let us assume, for instance, two binding energies of 0·50 and 0·77 ev. for Al_2O_3. If there are only a relatively small number of electrons having the smaller one, we expect to find, if n_0 is big, a straight line with a slope of about 0·25 ev., and if n_0 is small a straight line with ϵ equal to 0·385 ev. By measuring over a larger range of temperature, till a kink appears, this possibility could be distinguished from the one given in § 3. In the light of Meyer and Neldel's results both possibilities remain unsatisfactory.

§ 4. ACKNOWLEDGEMENT

I wish to express my thanks to Prof. Mott for suggesting the problem and for the many valuable discussions I have had with him.

---

* Not for the anomaly of ZnO, however.

REFERENCES

(1)(*a*) Gudden, B. *Sitzungsberichte der Physikalisch-medizinischen Sozietät zu Erlangen*, **62**, 289 (1930); cf. also for these and other questions, the review of:

(*b*) Gudden, B. *Ergebn. exakt. Naturw.* **13**, 223 (1934).

(2) Wagner, C. *Z. phys. Chem.* (*b*), **22**, 181 (1933).

(3) Wilson, A. H. *Proc. Roy. Soc.* A, **134**, 277 (1931); *Actual. Sci. Industr.* **82** (1933); *The theory of Metals* (Cambridge, 1936).

(4) Engelhard, E. *Ann. Phys., Lpz.*, (5), **17**, 501 (1933).

(5) Hartmann, W. *Z. Phys.* **102**, 709 (1936).

(6) Wagner, C. and Schottky, W. *Z. phys. Chem.* B, **11**, 163 (1930).

(7) Meyer, W. and Neldel, H. *Phys. Z.* **38**, 1014 (1938).

(8) Vogt, W. *Ann. Phys., Lpz.*, (5), **7**, 183 (1930).

(9) Schönwald, B. *Ann. Phys., Lpz.*, (5), **15**, 395 (1932).

(10) Kapp, G. and Treu, M. *Sitzungsberichte der Physikalisch-medizinischen Sozietät zu Erlangen*, **65**, 215 (1934).

(11) Jander, W. and Stamm, W. *Z. anorg. Chem.* **199**, 165 (1931).

(12) Baumbach, H. H. von and Wagner, C. *Z. phys. Chem.* B, **22**, 199 (1933).
(13) Ebert, F. (Cf. (12).)
(14) Mott, N. F. and Littleton, M. J. *Trans. Faraday Soc.* **34**, 485 (1938).
(15) Dünwald, H. and Wagner, C. *Z. phys. Chem.* B, **22**, 212 (1933).
Gundermann, J. and Wagner, C. *Z. phys. Chem.* B, **37**, 155 (1937).
(16) de Boer, J. H. and Geel, W. Ch. van. *Physica*, **2**, 286 (1935).
(17) Kröger, P. Unpublished; quoted by Gudden, *loc. cit.*
(18) Waibel, F. *Wiss. Veröff. Siemens-Kunz*, **10**, 65 (1931).
(19) Mott, N. F. *Proc. Roy. Soc.* A (in the Press).

COMPARISON OF COSMIC-RAY SHOWERS UNDERGROUND AND AT SEA-LEVEL

By D. H. FOLLETT, M.A., F.Inst.P.
Birkbeck College, London

Communicated by Prof. P. M. S. Blackett, F.R.S. 4 *February* 1939.
Read in title 12 *May* 1939

ABSTRACT. With assemblies of five and four counters used alternately, measurements of cosmic-ray shower-frequencies from various thicknesses of lead were carried out (*a*) at sea-level, under a thin roof and under a screen of bricks 1 m. thick, and (*b*) underground at a depth of 30 m. in a clay soil. Comparison of the results obtained with the two assemblies enables conclusions to be drawn regarding the size of the showers. The results at sea-level under a thin roof are in accordance with the predictions of the cascade theory of shower-production: underground the interpretation of the results is less unequivocal, but it is shown that when allowance is made for the spatial association of the rays incident on the counting assembly (due to the proximity of the roof of the tunnel in which the experiments were carried out) there is no reason to suppose that cascade processes of shower-production do not predominate there, as they do at sea-level. The results obtained at sea-level under a screen of bricks corroborate this conclusion.

§ 1. INTRODUCTION

In detail the results obtained by the various workers who have studied the shape of the curve connecting shower-frequency with thickness of absorbing material differ considerably. From a study of these differences in relation to the shower-counting dispositions employed, it is possible to draw conclusions regarding the mechanism of shower-production. For example, a number of workers have made measurements comparing the shower-frequencies observed when counting-assemblies employing different numbers of counters are used(1, 2, 3); in these investigations deductions were made regarding the number of particles in the showers, and they all lead to the conclusion that a cascade process must predominate, or at least play an important part, in shower-production.

In the work described below, comparisons similar to those made by HuChien-Shan and his collaborators(1) have been made at a depth of 30 m. of clay (where it is possible that the processes of shower-production may be different from those at sea-level), and are contrasted with results obtained at ground-level.

§ 2. APPARATUS

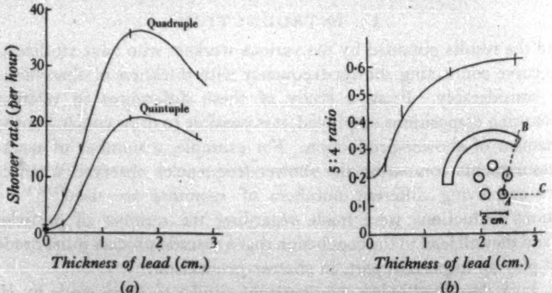
Throughout the work use was made of an assembly of five counters whose axes were arranged at the corners of a regular pentagon. Under each condition of shower-production, measurements of the shower-frequency were made alternately with all five counters and with one of the bottom pair disconnected from the circuit.

The counters were sealed in glass, the cathodes being of oxidized copper and the anodes of tungsten. The length of each was 30 cm. and the internal diameter of the glass envelope 1·6 cm. An orthodox coincidence-counting circuit was used, pentodes being employed in the first stage: a telephone counter, operated by a thyratron, recorded the coincidences. In most of the work, an arrangement was used whereby at regular intervals one of the pentodes was automatically cut out of the circuit: every twenty-four hour run was thus divided into alternate periods comprising respectively four hours of quintuple-coincidence counting and two hours of quadruple-coincidence counting; for each separate quintuple-counting period the counts were recorded on a separate telephone counter, thus affording a check on the steady running of the set. The alternation between quintuple and quadruple counting eliminates the effect of any small secular changes which might occur in the characteristics of the set (for example, as a result of fluctuations in the counter potentials, which were provided by batteries).

§ 3. RESULTS

Measurements of the quintuple-coincidence and quadruple-coincidence counting rates were carried out under four different sets of conditions, as detailed below.

Figure 1 (*a*) shows the quintuple and quadruple counting rates at ground-level for thicknesses of lead up to 2·7 cm., the apparatus being set up in a small building with thin walls and roof. In figure 1 (*b*) is shown a scale diagram of the disposition of the counters and lead used. The counter marked *A* is the one which was left out when quadruple rates were being counted. The lead screen is made in the form of an incomplete hemi-cylinder. This ensures that the angular spread of all the showers detected shall be the same; for if the hemi-cylinder is supposed to be complete, and if showers starting in the lead between the lines *BA* and *AC* are considered, those detected by the quadruple assembly would have a rather smaller spread than those detected by the quintuple assembly; and since the size of a shower and its angular spread may be related, this might lead to a certain vitiation of the

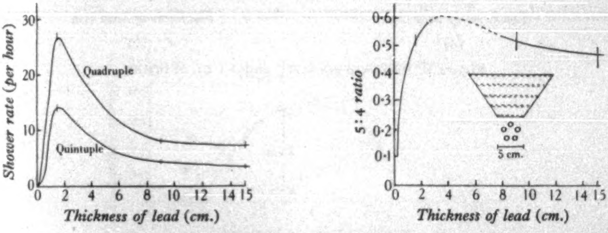


(*a*) (*b*)

Figure 1. Results at sea-level, under a thin roof.

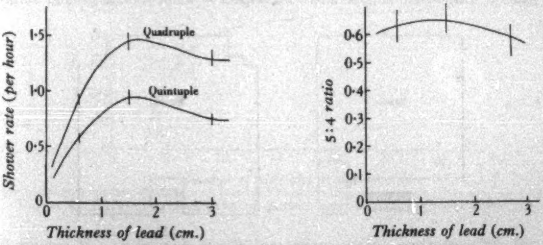
results. Figure 1 (*b*) shows the ratio of the quintuple to the quadruple counts. These ratios will be referred to henceforward as "5 : 4 ratios", for brevity.

Figure 2 (*a*) shows quintuple and quadruple counting rates extended to 10 and 15 cm. of lead. For these thicknesses the disposition of the lead was different from that for the earlier readings, for the sake of economy in lead; and in order that the geometric conditions should be comparable for all three thicknesses the lead was arranged in separated layers for the 15-cm. and 10-cm. readings, the separation of the upper and lower boundaries of the lead remaining constant at all three thicknesses; the dotted lines in the figure show how it was arranged for the reading at 15 cm. Figure 2 (*b*) shows the 5 : 4 ratios.



(*a*) (*b*)

Figure 2. Further results at sea-level under a thin roof.

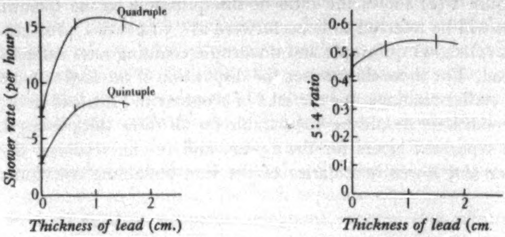


(*a*) (*b*)

Figure 3. Underground results.

Figures 3 (*a*) and 3 (*b*) show the quintuple and quadruple rates and their ratios measured underground at a depth of 30 m., under a clay soil having a density of about 2, in Holborn Underground station; the position of the apparatus with respect to the walls of the tunnel is indicated by *X* in figure 5. The arrangement of the counters and lead was the same as in the first set of readings.

Figures 4 (*a*) and 4 (*b*) show the results obtained with counters and lead arranged under a screen of bricks about 1 m. thick. The disposition of the bricks with respect to the counting system is shown in two elevations in figures 6 (*a*) and (*b*).



(a) (b)

Figure 4. Results at sea-level, under 1 m. of bricks.

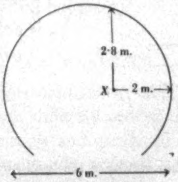
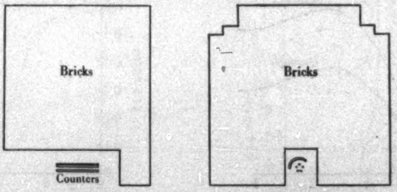


Figure 5. Disposition of apparatus with respect to walls of underground tunnel.



(a) (b)

Figure 6. Disposition of brick screen and counter-assembly.

§4. DISCUSSION OF RESULTS

General considerations. C. G. Montgomery and D. D. Montgomery(4) have derived formulae for the probability of detection of a shower of N particles by an n-fold coincidence-counting assembly, by making certain rather specialized simplifying assumptions. The application of these expressions to the counter-assemblies here used is not entirely straightforward, as the effective values of n in the two cases is not clear; but assuming it to be 4 and 3 respectively in the quintuple and quadruple assemblies, the curves of figure 7 are derived. Figure 8 shows the ratios of these probabilities plotted against N. These curves can be expected to represent only very approximately the true probabilities; but they are of assistance in affording a rough quantitative basis for the interpretation of the results.

The following general considerations regarding the relations between the size of showers and the observed 5 : 4 ratios are of importance. (*a*) If all the showers from some arrangement of absorber were of the same size, then the number in each case could be deduced from the observed 5 : 4 ratio. It might be thought also that, even if the showers had a wide distribution in size, the 5 : 4 ratios in these experiments would be a measure of the average size of the showers: this, however, is not true, as the form of the {size, frequency} distribution curve will affect the 5 : 4 ratio. (*b*) The calculated 5 : 4 ratio increases rapidly at first with increasing N and even when $N=10$ is not much less than unity: consequently the observed value of the 5 : 4 ratio is influenced only by variations in the proportions of the smaller showers, for which N is less than about 10; and it will be particularly sensitive to changes in the proportions of the smallest showers, for which N is less than about 6, over which range the 5 : 4 ratio increases rapidly. (*c*) All experiment and theory at present indicate that under all conditions of shower-production there is a big proportion of showers containing a few particles only, so that we may expect to find that the observed 5 : 4 ratio always remains well below unity.

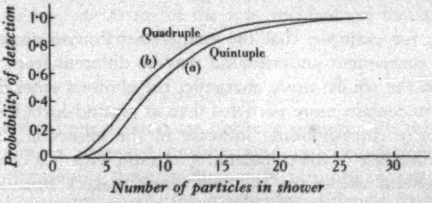


Figure 7. Probabilities of shower-detection by the two assemblies.

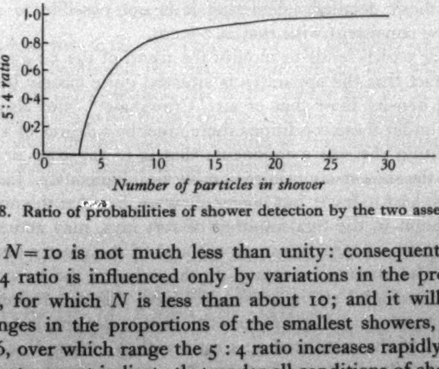


Figure 8. Ratio of probabilities of shower detection by the two assemblies.

Ground-level results under a thin roof. The steady rise of the observed 5 : 4 ratio with increasing thickness of lead, figure 1, indicating the progressive decrease of the proportion of very small showers, is in accordance with what one would expect from the cascade theory. The value of the 5 : 4 ratio for large thicknesses of lead indicates the presence of a large proportion of small showers (i.e. showers for which

N is less than about 6), which is in accordance with the supposition that the showers emerging from a large thickness of lead are due not to electrons, positrons and photons incident on the lead, but to secondary electrons and photons produced in the lead by the mesotrons constituting the penetrating component of the cosmic radiation.

Underground results. At ground-level, where cascade processes of shower-production are known to predominate, the 5 : 4 ratio increases rapidly with increasing thickness of lead up to 1·5 cm., figure 1, while underground the ratio is much more nearly constant, figure 3. This might suggest that cascade processes do not predominate underground but, if present at all, are masked by non-cascade processes such as those suggested by Heisenberg and others. There are, however, other more likely explanations.

It is possible, for example, that the energy-distribution among the rays constituting the soft component underground may be different from that at sea-level: if the rays are, on the whole, more energetic, the showers emerging from 0·6 cm. of lead will tend to contain more particles than at ground-level for the same thickness, thus causing a corresponding increase in the observed 5 : 4 ratio; and in view of the rapid increase of the calculated 5 : 4 ratio with N a comparatively small increase in the general size of the showers would make a considerable difference to the observed 5 : 4 ratio. One would expect also an increase in the 5 : 4 ratio for 1·5 cm. compared with that at sea-level; a small increase does indeed appear to exist, but without detailed calculation it is not possible to say whether this increase would be consistent with that at 0·6 cm.

Another factor which tends to modify the trend of the 5 : 4 ratio underground arises from the fact that the apparatus is situated quite closely beneath a medium of much greater density than that of air. Crawshaw(5), and also Jànossy(6), have pointed out that under these conditions there must be a degree of association among the rays constituting the soft component: that is to say, two or more rays must frequently strike the shower-counting assembly simultaneously. Jànossy has pointed out that this must lead to fewer but bigger showers, but that the number of showers recorded, in relation to the total number of soft rays, may actually be much the same as if there were no association, on account of the greater probability of their being detected. The observed increases in the value of the 5 : 4 ratios relative to the values at ground-level are in accordance with this conclusion; and the much greater increase at 0·6 cm. as compared with that at 1·5 cm. is to be explained by the rapid increase of the calculated 5 : 4 ratio with N, when N is small.

Ground-level results under a brick screen. The 5 : 4 ratios obtained under a screen of bricks at ground-level, figure 4, show a similar trend to those obtained underground. The brick screen is probably thick enough to filter out all the soft radiation incident on it, so that all the soft radiation incident on the lead has probably originated in the bricks secondarily to the penetrating component of the incident radiation. Probably, therefore, conditions are roughly comparable to those underground, except that the counter system is much closer to the high-density material than it is underground: this leads to a higher degree of association of the rays incident on the counter-assembly, as is indicated by the relatively large number of showers when there is no lead over the counters. The observed values for the 5 : 4 ratios are actually a little lower than those found underground, whereas the greater degree of association would lead one to expect higher values: the differences are barely significant, however. If the difference is real it may perhaps be explained in terms of the very small showers which, under conditions of non-association, are recorded either not at all or in very small proportions: with the very high degree of association that must exist under the brick screen, composite showers of these particles are probably large enough to be recorded in sufficient quantities to influence the observed 5 : 4 ratio, while remaining small enough to make the ratio smaller than it is underground. The difference may also be accounted for by a difference in the energy spectrum of the rays incident on the lead, as compared with the underground energy spectrum.

§ 5. CONCLUSION

The results of this investigation are not inconsistent with the assumption that the shower phenomena observed underground can be adequately explained in terms of cascade processes: there is apparently no need to assume the existence underground of any shower-producing mechanism different from those obtaining at sea-level.

§ 6. ACKNOWLEDGEMENTS

I wish to express my thanks to Prof. P. M. S. Blackett, who suggested this investigation, for his continued interest, encouragement and advice. My thanks are also due to Dr H. J. J. Braddick, Mr E. P. George and Dr L. Jànossy for help received from discussions with them; to the London Passenger Transport Board for permission to work on the disused platform in Holborn Underground Station; and to the station staff for assistance in many ways.

REFERENCES

(1) HuChienShan, Kisilbasch, B. B. and Ketiladge, D. *Proc. Roy. Soc.* A, **161**, 95 (1937).
(2) Froman, D. K. and Stearns, J. C. *Phys. Rev.* **52**, 382 (1937).
(3) Geiger, H. and Heyden, M. *Z. Phys.* **110**, 310 (1938).
(4) Montgomery, C. G. and Montgomery, D. D. *Phys. Rev.* **48**, 786 (1935); *J. Franklin Inst.* **221**, 59 (1936).
(5) Crawshaw, J. D. *Proc. Phys. Soc.* **50**, 783 (1938).
(6) Jànossy, L. *Proc. Roy. Soc.* A, **167**, 499 (1938).

THE STRUCTURE OF PROTEINS

By IRVING LANGMUIR

Address delivered 20 *December* 1938

§ 1. INTRODUCTION

THE proteins form one of the essential constituents of all living matter. A huge number of different varieties of protein exist in animals and plants. Apparently new kinds of proteins can be made at will by injecting into any animal a given protein foreign to that animal. The animal then develops an antibody that reacts with the foreign protein but is itself a protein. It looks as if we can force animals to make for us almost as many new kinds of proteins as the product of the number of animals and the number of proteins that already exist. Thus one of the wonderful characteristics of the proteins is their enormous number and therefore complexity.

When we are dealing with a substance as closely related to life as a protein is, we may expect to find structures that we have not met with before in organic chemistry.

To understand the structure of an animal, we must not only study its anatomy by microscopical and chemical examination of all parts of its body, but we must go into its physiology and biochemistry and particularly into the history of its evolutionary development, during the course of ages, through mutations and natural selection.

Similarly to understand the structure of a protein we shall need to consider not only its chemical composition and its physical properties, which correspond to its anatomy, but its specific biological reactions and functions. We must recognize also that the proteins are substances that have come into existence through the processes of evolution. Stanley has found that the virus of the tobacco mosaic disease, which he has shown to be a protein, can reproduce itself in the appropriate environment in the tobacco plant very much as if it were a living organism, and that it is capable of undergoing mutations which breed true. Thus we see, at least for a complicated protein of this kind (its molecular weight is about 50,000,000), that natural selection can act to preserve the mutant protein. Simpler proteins have been found to be formed by autocatalytic reactions from precursors; so here too we have evidence that proteins reproduce themselves in the proper environment, namely in the presence of the precursor.

Thus the proteins possess many of the attributes which have previously been thought to be typical of living beings. We should therefore not be surprised to find in the proteins features of structure which are quite new among organic molecules and which involve an organization and differentiation of parts that shows some approach to that of the simplest of living organisms.

§ 2. THE FUNDAMENTAL DATA

Of course, at present, our knowledge of the evolution of proteins is practically nil, so that we cannot yet trace the development of proteins from their still simpler beginnings. We have to be content with theories of protein structure based upon the known facts regarding these substances. From the enormous number of data that have been collected by chemists, physicists and biologists we must select those that seem particularly relevant to questions of structure. I have twelve chosen groups of data that seem to me to be of the most significance.

(1) *The amino-acid composition of proteins.* About twenty-five varieties of amino-acid molecules, $NH_2—CHR—COOH$, represented by different side-chains or R groups, have been found among the degradation products of proteins. The chemist has shown that proteins are built up of such molecules, joined together by the elimination of water, and has suggested that proteins consist of polypeptide chains

$$NH_2—CHR—CO—(NH—CHR—CO)_n—NH—CHR—COOH$$

in which the amino-acid residue functions as a two-armed unit

$$—(NH—CHR—CO)— \qquad ......(A).$$

(2) *The amino-acids in proteins are of α type.* The molecules thus have a common backbone $C—C_\alpha—N$, and none of them have backbones such as $C—C_\alpha—C_\beta—N$, $C—C_\alpha—C_\beta—C_\gamma—N$.

(3) *The laevo configuration of* C_α *in the amino acids in proteins.* The C_α atom in amino acids can be right-handed or left-handed. However, all the amino acids formed by the degradation of proteins are of laevo type.

(4) *The valency angles and distances apart of carbon and nitrogen atoms.* Crystallographical data have shown that carbon and nitrogen when single bonded in general have tetrahedral valency angles, and that the distance between single bonded carbon atoms and between single bonded carbon and nitrogen atoms is about 1·5 A.

(5) *There exist globular proteins which have definite molecular weights discretely arranged.* It has been found by Svedberg and co-workers that many proteins have definite molecular weights and that the molecular weights fall into a number of discretely arranged molecular-weight classes. Each molecular-weight class has a considerable spread. Svedberg finds a class at about 18,000, another at about 36,000, and a number of other molecular-weight classes, practically all of which are multiples of about 18,000. Evidence has also been found which indicates that these proteins are in some sense globular and that some of them, e.g. insulin, are approximately spherical. These researches seem to indicate that these proteins are definite compounds, and there is now a general consensus of opinion among protein chemists that each of these proteins is a definite chemical individual, all of whose molecules are alike.

Now that is quite a different state of affairs from what would be expected from long-chain molecules. The chemist knows how to make many kinds of polymers which are long chains, but these are quite indefinite in length. According to the way in which they are made, the lengths have different average values. Such polymers are not definite chemical individuals, and as such are to be contrasted with the proteins.

(6) *The numbers of individual amino-acid residues in proteins are in many cases powers of* 2 *and* 3. In trying to find out as much as he can about the constitution of proteins, the organic chemist has met with a certain amount of success. For instance, some of the latest results obtained by Max Bergmann in New York may be cited. He has determined the numbers of various types of amino-acid residues in certain proteins and concludes that these numbers are always powers of 2 and 3. In the particular case of egg albumin, whose molecular weight is about 36,000, he concludes that each individual molecule contains 288 residues. If we consider the many thousand different ways in which a molecule of this order of molecular weight can be made up of different proportions of the available amino acids, these are remarkable results. Other proteins also have been considered in this way, and Bergmann has suggested that there are 2 × 288 residues in the molecules of ox haemoglobin and ox fibrin. Facts of this kind undoubtedly tell us something about the structure of a protein molecule.

(7) *The globular proteins have extremely specific properties.* Perhaps the most striking of all facts about the globular proteins is their high degree of specificity. Thus haemoglobin in each different type of animal seems to differ in spectrum and in detail of its behaviour with oxygen. Immuno-chemistry in general draws the same picture of globular proteins as highly organized structures with extremely delicate specificities, as evidenced by antigen-antibody reactions.

Not all of the specific reactions involve interactions between proteins; some proteins catalyse decompositions of very simple substances. For example, urease changes urea into ammonia, catalase liberates oxygen from hydrogen peroxide and other enzymes act on carbohydrates, and so on.

(8) *The denaturation of the globular proteins.* The globular proteins have an extraordinary property, namely the capacity to become denatured. As far as I know, that term is not used to describe the properties of any substance except proteins. For example, pepsin in a concentration of one part in 50,000 or so causes milk to clot, and this property is in fact used as a measure of the activity of the pepsin. Now this specific property (and other specific properties in other cases) disappears quite abruptly under certain treatments. The activity of a solution of pepsin can be reduced and finally destroyed by shaking in a partly filled bottle. It can be destroyed by heating to 65° for a few minutes and in many other ways. Denaturation usually renders a soluble protein insoluble; but in the presence of sufficient acid or alkali it may remain soluble, and the molecular weight under such conditions has been found to be substantially the same as that of the native protein.

There exist, particularly in animals, large numbers of substances which have the general chemical composition of proteins, and are therefore called proteins, but do not form true molecular solutions and are thus to be contrasted with the globular proteins. Examples are furnished by the extra-cellular proteins, such as the collagens, keratins (hair, wool, scales, feathers, etc.), and silk proteins, which have, or by stretching can be made to have, a fibrous structure. These are probably all formed in the animal body by the denaturation of globular proteins. Denaturation appears to be a very remarkable property of the globular proteins. We find nothing like it with substances having molecules of ordinary size, or in the long-chain molecules of high molecular weight, nor would we expect to do so.

(9) *Many globular proteins, but no denatured proteins, have the power to crystallize.* Many of the globular proteins form crystals in spite of the size of their molecules, but it appears that no denatured protein has ever been obtained in crystalline form. The crystals of the globular proteins are remarkable in that they are of high symmetry, often cubic or hexagonal, whereas most organic substances, except those having very small molecules, give crystals of low symmetry such as monoclinic or triclinic(2–4).

The ready formation of highly symmetrical crystals from pure globular proteins and the absence of crystals from denatured proteins indicates that the molecules of the globular proteins are themselves structures possessing a high degree of symmetry, and that this structure is destroyed by denaturation.

(10) *Many of the globular proteins are very soluble in water or salt solutions.* The solubility or insolubility depends largely upon the presence of and availability of hydrophilic and hydrophobic groups. Only those groups that form the surface of the molecule determine the solubility. Evidently, then, the solubility and its modification by various agents should give information regarding the distribution and arrangement of these groups in the molecules.

(11) *Many proteins when in solution spontaneously form monolayers which are extremely insoluble.* This insolubility indicates the presence of hydrophobic groups which are made available by the spreading process. A study of the formation and properties of monolayers given by globular and by denatured proteins thus throws light on the structure of the globular proteins and the modifications produced by denaturation(5–7).

(12) *A number of globular proteins give* X-*ray photographs.* X-ray photographs have been obtained in recent years from crystals of pepsin, insulin, haemoglobin, chymotrypsin, lactoglobulin, tobacco-seed globulin and other proteins. Wet crystals, which often contain over 50 per cent of water, give far better photographs than dry crystals. In some cases the data permit the measurement of many hundreds of reflection intensities. The X-ray data indicate that the crystals are not mere lattice-like arrangements of ill-defined molecules, but that the molecules themselves possess high symmetry and a fine-grained internal structure implying an organization or architecture quite unique for molecules of such size. X-ray methods, in the future, should provide data by which all details of protein structure may ultimately be found. At present the difficulties of interpretation of such data are extremely great, so that the X-ray data now serve only as a test of theories of structure derived from other sources.

The twelve points summarized briefly above give an outline of the available facts relating to the proteins, and any theory of protein structure, or any views regarding protein structure, should be consistent with as many of these phenomena as possible.

§ 3. THE GLOBULAR PROTEINS

Although the globular proteins differ from one another very greatly in many properties, they possess in common some remarkable features which suggest that all the globular proteins have a highly organized type of structure that distinguishes them from all other substances. How are we to find this characteristic structure of the molecules of the globular proteins?

In general, the structure of organic molecules has been investigated by organic chemists, and in this they have been very successful. Many years ago, while I was discussing with Nils Bohr the extraordinary scientific achievements of the nineteenth century, he expressed his admiration of the great accomplishments of the organic chemists in deriving the structures of aromatic compounds and stereo-isomers. I think physicists realize only imperfectly the magnificent feat which the chemists performed in finding so many structures, without any of the tools which the physicist of to-day has come to think essential.

These chemical methods, often thought of as involving chemical intuition, are particularly applicable to cases which are so complicated that the modern direct methods of the physicist are not available. Thus we should expect that the cumulative evidence obtained from a study of the general physical and chemical properties of proteins might serve as a foundation for a theory of the characteristic structural features of the proteins, even if it is as yet premature to attempt to derive the complete structure of any single protein.

During the gradual development of organic chemistry, many specific reactions or tests for the presence of certain groups of atoms or types of binding between atoms have been found, but none of these have been shown to be characteristic of the globular proteins. This failure appears to be associated with the extreme ease with which the native proteins are denatured. It is evident that the really characteristic features of the globular proteins are destroyed by most of the treatments to which the chemist subjects the proteins during his study of them. For example, in deriving the amino-acid constitution of the proteins the usual initial procedure is to boil the protein for a long time with acid. Since some proteins in solution are denatured by heating for a few minutes at 65°, this treatment is hardly likely to yield much detailed information about the way in which the amino-acid residues were joined in the original protein. It is as though, to learn about the architecture of Ypres, you were first to let an army bombard it, and then to examine the ruins.

The situation is somewhat analogous to that met with in quantum physics, where the experimental conditions required for observing the position of an electron modify the velocity and so prevent an accurate measurement of the original velocity.

Although the detailed techniques of orthodox organic chemistry may thus throw little light upon the characteristic structure of the native proteins, the general inductive methods of reasoning which were responsible for the early development of organic chemistry are admirably adapted for an attack on the problem on the basis of the known general physical and chemical properties as outlined in the twelve points that have been listed.

When the chemist found substances consisting of polypeptide chains among the degradation products of proteins, he made the natural inference that proteins in general have polypeptide structures. According to this picture the protein-building unit is the two-armed amino-acid residue which can form unbranched chains of any length. The proteins should thus differ from one another only in the kinds of amino-acid residues present and in their linear sequence. The number of possible different ways in which twenty kinds of residues could be arranged in single chains containing two or three hundred such residues is of course enormous and could perhaps account for the large number of proteins that can be formed.

Such a structure may account satisfactorily for the properties of the fibrous proteins and many of the products formed by the denaturation or degradation of the globular proteins. A simple polypeptide chain theory, however, does not appear to be compatible with the properties of the globular proteins as given in items 5, 6, 7, 8, 9, and 10 of our list.

Let us consider for a moment some of the properties that chemists have found to be characteristic of organic molecules consisting of chains of atoms such as carbon and nitrogen. When the molecules are free, as for example when they are in solution, the parts of the molecule are nearly free to rotate about every axis corresponding to the bonds connecting adjacent atoms that form the chain. Such molecules should therefore not have characteristic shapes. The simple crystalline form (item 9 § 2) given by globular proteins shows, however, that their molecules have definite and symmetrical shapes.

The chemist knows also that the reactivities and other properties associated with a group attached to one of the links of a chain are modified by the presence of certain groups attached to neighbouring links. Thus chemical effects can be transmitted along the chain. For example, consider derivatives of propionic acid $CH_2R—CH_2—COOH$. For propionic acid itself, where R represents a hydrogen atom, the dissociation constant, which measures the ease with which a hydrogen ion escapes from the $—COOH$ group, has the value $1{\cdot}4 \times 10^{-5}$. With β-chlorpropionic acid, in which R represents a chlorine atom, the dissociation constant has increased to $8{\cdot}6 \times 10^{-5}$. Some effect has thus been transmitted along the chain from the β carbon atom to the carboxyl group. An examination of similar data for a series of chlorine-substituted fatty acids(8), in which the chlorine is attached to different carbon atoms of the chain, has shown that the effect transmitted along the chain decreases exponentially as the length of the transmission path increases, each intervening carbon atom decreasing the effect in the ratio 2·7 : 1. The effects directly transmissible along the backbone of a polypeptide chain should thus extend only to distances of one or two residue-lengths.

The specific properties of the globular proteins, however, prove that effects are transmitted throughout practically the whole molecule. In the case of interactions between proteins, such as antigen-antibody reactions, one might assume that the specificity depends upon the lengthwise association or fitting together of two linear polypeptides much as two chromosomes, one from the male and the other from the female germ cell, join along their lengths. Thus the linear sequence of the amino acids in a polypeptide might have a significance similar to that of the genes in the chromosomes. But this hypothesis cannot explain the specific reactions of urease, catalase or haemoglobin where single protein molecules react with simple substances such as urea, hydrogen peroxide or oxygen. The fact that the absorption spectrum and oxygen affinity of the haematin in the haemoglobin molecule depends upon the specific properties of the globin to which it is bound proves the long-range transmission of chemical effects within the framework of single globin molecules.

The properties of crystals of straight chain hydrocarbons and their derivatives, such as the normal fatty acids and alcohols, are frequently considerably modified by the addition of one —CH_2-link to the length of the chain. This, however, does not indicate an effect transmitted directly through the chain from one end to the other, for it results from the fact that the crystallizability depends upon the symmetry of the molecule as a whole. Thus the melting points and other properties of the crystals frequently show alternating positive and negative increments upon the addition of successive —CH_2 groups, while the properties of the liquids show no such irregularities. Evidently then the crystal forms given by compounds consisting of simple polypeptide chains (if these are short enough to give crystals at all) might be very sensitive to the particular amino acids present and to their sequence, but such sensitivity would not appear in the properties of solutions of these substances.

Chemists have long known that two substituents in the benzene molecule have a particularly large effect on one another. It is now known that this is an example of electronic resonance that occurs when adjacent double and single bonds can give a kind of tautomerism between two possible configurations. Resonance tends to give an increased stability to a structure and greatly increases its power to transmit chemical effects between groups connected by the resonating atoms. Hydrogen atoms which act as bridges between two oxygen atoms (hydrogen bonds) also give resonance, for the single bond of the hydrogen atom in the classical valence theory may be thought of as alternating in position between the hydrogen atom and each of the two oxygen atoms.

The extraordinarily great transmission of chemical effects throughout the molecules of the globular proteins thus seems to demand a structure whose parts are bound by resonating atoms. The simple polypeptide chain with its singly bonded carbon and nitrogen atoms should show no resonance and therefore cannot account for the properties of the globular proteins.

The discrete molecular weights of the globular proteins (item 5) and the occurrence of only powers of 2 and 3 in the number of particular residues (item 6) seem to be incomprehensible on the basis of a structure consisting of a simple polypeptide chain. Bergmann and Niemann have suggested (1b) that each kind of residue occurs at regular intervals along the length of the chain as determined by a kind of resonance of the chain as a whole, nodes and loops being formed at regular intervals. In view of the absence of resonating atoms in the simple polypeptide chain, and of the fact that the differing composition of the residues in the successive intervals between the residues of any one kind would make the intervals unequal, we must reject this as a purely *ad hoc* hypothesis.

The phenomena of denaturation (item 8) also seem incompatible with a simple polypeptide structure for the globular proteins. The treatments required for the milder forms of denaturation are incapable of breaking apart the polypeptide chain or causing an alteration of the sequence of the residues in such a chain. How therefore could they have such a profound effect on the specific properties which are assumed to depend on this sequence?

The high solubility in water of many globular proteins (item 10) proves that the *surfaces* of these molecules are almost wholly hydrophilic, yet the extreme insolubility of the monolayers formed from these proteins (item 11) proves the presence of a large proportion of hydrophobic groups in the molecules(6). Such a change in the availability of the hydrophobic groups indicates that there is a radical difference in structure between the molecule of the globular protein and the monolayer formed from it. If both were to consist of simple polypeptide chains, with free rotation about the bonds between the links, no such difference in properties is conceivable.

Many protein chemists, while maintaining that a single polypeptide chain forms the backbone of the protein molecule, now recognize that there are many reasons for believing that the chains of separate molecules in protein fibres and different parts of the same chain within single molecules are interconnected by cross linkages. These are usually assumed to involve the side chains, particularly those containing polar groups, or to consist of disulphide bridges formed from two cysteine residues. Such cross linkages, although they may often exist and account for some properties of proteins, would give no resonance and could not explain the transmission of chemical effects that must be responsible for the specificities of the globular proteins (item 7). The number of such linkages within any one molecule should be relatively small and very greatly dependent upon the presence of particular residues, and thus would not give the highly organized characteristic structure of the globular proteins needed to account for the properties listed in items 5 to 12.

§ 4. MANY-ARMED UNITS

A few years ago Dr Wrinch attempted to study the structure of chromosomes(9). Realizing that this problem was insoluble without a knowledge of the structure of the substances of which they are composed, she proceeded to analyse the available data relating to the structure of proteins. Starting with the datum that the proteins are condensations of amino-acid molecules (item 1), using the tetrahedral valency angles for carbon and nitrogen atoms (item 4), taking a definite mean value a (say 1·5 A.) for interatomic carbon-carbon and carbon-nitrogen distances, and specifically using the datum that all the amino acids are of α type (item 2), she attempted to work out possible structures for protein molecules, correlating these sets of facts into a single scheme(10).

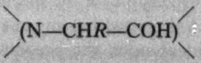
In a simple polypeptide chain theory no particular significance is to be attached to the fact that the constituents of proteins have the common backbone unit $-C-C_\alpha-N-$ and never for example a unit $-C-C_\alpha-C_\beta-N-$, yet so striking a fact must be fundamental for the structure of proteins. The two-armed building unit (*A*),

$$-(\mathrm{NH}-\mathrm{CH}R-\mathrm{CO})- \qquad \ldots\ldots(A)$$

however, seems to require a simple linear arrangement of the units giving a straight-chain polypeptide. A building unit having greater combining capacity, such as one with four arms, would permit of the formation of structures covering surfaces or filling volumes which would be capable of having a much higher degree of organization than a mere linear array of units.

The need for such a many-armed unit is also strongly indicated, I believe, by the facts of genetics. The chromosomes in a fertilized germ cell must contain the complete specifications of the adult animal that develops from it. When one considers, for example, the enormous complexity of the arrangement of cells in the human brain, and the inherited instincts and capabilities that they carry, one cannot but marvel that the specifications for so many details can be packed into a volume as small as that occupied by the chromosomes. Certainly to solve such a problem Nature has needed to utilize all available structural resources in the geometrical arrangements of the atoms within the proteins that form the chromosomes. A restriction to a mere linear arrangement in an unbranched chain would enormously limit the structural possibilities. It seems highly probable that Nature has taken full advantage of the extra degree of freedom provided by the 2-dimensional or 3-dimensional arrangement of units permitted by a many-armed unit.

Adopting a suggestion of F. C. Frank(11), Dr Wrinch has postulated the four-armed building unit (*B*):



$$\rangle(\mathrm{N}-\mathrm{CH}R-\mathrm{COH})\langle \qquad \ldots\ldots(B)$$

obtained from the two-armed unit (*A*) by the transfer of a hydrogen atom or proton from the nitrogen of one residue to the adjacent —CO group in the next residue. To realize the vastly increased number of possible structures permitted by the four-armed unit as compared with those given by a two-armed unit, one only needs to contrast the chemistry of the tetravalent element carbon with that of divalent elements like oxygen.

Dr Wrinch then proceeded to investigate mathematically the possible arrangements of these four-armed units and to determine whether any of the resulting structures possess features that correspond to the known properties of proteins. The cyclol theory is concerned with the implications and consequences of this postulate of the four-armed unit. Since this unit has the composition of an α-amino acid residue and the asymmetric carbon atom in it may be assumed to have the laevo configuration, structures given by the cyclol theory automatically conform to items 1, 2, and 3 of our list. Dr Wrinch has confined herself to structures in which the tetrahedral valency angles characteristic of the carbon and nitrogen atoms are preserved. Thus the structures also conform to item 4.

§ 5. THE CYCLOL FABRIC

Just as the two-armed units (*A*) can give an endless chain, so the four-armed units (*B*) can give a 2-dimensional array or plane fabric that can extend indefinitely. The lace-like pattern of this cyclol fabric is shown in figure 1. It is seen that the carbon and nitrogen atoms that form the skeleton of the fabric form six-membered rings that are of two kinds. There are diazine rings which contain two nitrogen atoms, two CH*R* groups and two COH groups. There are also triazine rings which contain three nitrogens and three COH groups. Each diazine ring links together two triazines, while each triazine is joined to three diazines. The fabric as a whole has trigonal symmetry about any triazine ring. The three —OH groups attached to the three carbon atoms of any triazine ring lie on the same side of the plane of the fabric, but in three neighbouring triazines (beyond the adjacent diazines) the oxygens lie on the opposite side of the plane.

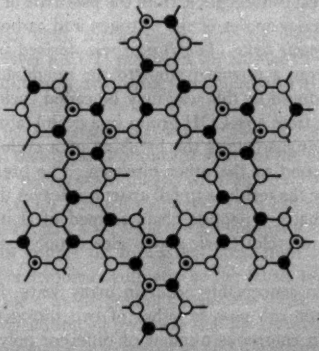


Figure 1. A fragment of the cyclol fabric. The median plane of the lamina is the plane of the paper. The lamina has its "front" surface above and its "back" surface below the paper. ● = N. ○ = C(OH), hydroxyl upwards. ⊙ = C(OH), hydroxyl downwards. ○— = CH*R*, direction of side chain initially outwards. ○- = CH*R*, direction of side chain initially upwards.

A characteristic feature of the cyclol fabric shown in figure 1 is the presence of lacunae or openings around which there are six triazine rings at the corners of a hexagon and six diazine rings that form the sides of the hexagon. The side chains that are attached to the carbon atoms on the near side of each diazine ring are arranged around the lacuna with trigonal symmetry. Three of these side chains are bound to the carbon atoms of the fabric by bonds that rise vertically upward from the median plane of the fabric, while the other three are attached by bonds that form an angle of 71° with the normal to the plane. Thus the bonds that hold the side chains all lie on one side of the median plane. The fabric is thus dorsiventral, i.e. its two sides are different. This characteristic results from the fact that all the amino acids have the same configuration (item 3). This puzzling uniformity thus receives a simple interpretation on the cyclol theory.

The fact that only α-amino acids are present in proteins (item 2) also becomes understandable on the basis of the cyclol theory, for only by having building units of equal length, and in fact only with three atoms in the chain, can a symmetrical plane fabric structure such as the cyclol fabric be constructed.

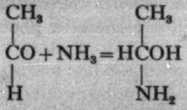
Other types of fabric. It should be noted that the cyclol fabric, figure 1, contains only *B* units; Dr Wrinch has shown that by using both *A* and *B* units it is possible to construct a large number of other fabrics which, however, all have larger lacunae and therefore a less compact structure(10b). Although she has made many attempts she has not yet succeeded in constructing any plane fabric, other than that shown in figure 1, using only *B* units. It seems therefore improbable that any other fabrics of this type exist. The cyclol fabric thus seems to be derivable automatically, as a definite mathematical consequence, from the postulate of the four-armed *B* unit and the tetrahedral valency angles of the nitrogen and carbon atoms.

Cyclol bonds and peptide links. The *B* unit was derived from the *A* unit by the transfer of a proton from a nitrogen atom to the neighbouring CO group. The pair of new bonds produced in this way may be called cyclol bonds. Conversely, the passage of a proton from any COH group in a cyclol fabric to either one of the two adjacent nitrogen atoms causes the breaking of a cyclol bond but leaves the resulting CO group bound to the other nitrogen. The breaking of cyclol bonds therefore does not allow the residues to fall apart but leaves them bound by peptide linkages as in the normal polypeptide chain formed of *A* units. The breaking of all the cyclol bonds in a cyclol fabric leaves closed polypeptide chains consisting only of *A* units.

This breakdown can conceivably occur in many ways, for in the breaking of one cyclol bond a proton can pass to either of two nitrogen atoms(12). It might seem, therefore, that an enormous number of different polypeptide chains would result from the degradation of a single protein. It is more probable, however, that the strength of any cyclol bond depends on the nature of the neighbouring side chains, and thus the breakdown of the fabric may be expected to follow a preferred path giving a very limited and definite set of closed chain polypeptides.

Chemists have raised several objections to the cyclol bond, upon which the cyclol theory is based. Some chemists assert that a bond of this type has never been found in chemical compounds and therefore must be too unstable to exist. One must remember, however, that the substances whose structures the chemists determine are not often broken down by such mild treatments as those that cause denaturation of proteins. In fact, the cyclol bond, if it is to explain the structure of the globular proteins, must be so unstable that the chemist should not have found it by the methods he uses. There have been some attempts to calculate the energy of formation of cyclol bonds and so to prove that they are too unstable to be formed. The theoretical basis of such calculations and the data upon which they rest are as yet too uncertain to be of much value, especially as they fail to take into account several stabilizing factors that we shall discuss presently.

Another objection often raised is that covalent bonds between carbon and nitrogen, such as those that act both between the *A* units in the polypeptide chains and between the *B* units in the cyclol fabric, involve too much energy to permit the ready formation and breakdown of the cyclol fabric. The formation of the cyclol bond, however, involves not merely new bonds between carbon and nitrogen but also causes an NH group to lose its hydrogen and changes a CO group into a COH, changes whose energies may compensate for that of the carbon-nitrogen bond. A definite answer to this type of objection is given by the observed reversible formation of aldehyde-ammonia, by the reaction:



$$\begin{array}{c} CH_3 \\ | \\ CO \\ | \\ H \end{array} + NH_3 = \begin{array}{c} CH_3 \\ | \\ HCOH \\ | \\ NH_2 \end{array}$$

Here the nitrogen atom from the ammonia attaches itself directly to the carbon of the CO group and gives a COH just as in the formation of the cyclol bond. The reaction gives an equilibrium at ordinary temperatures so that it must be easily reversible.

Folding of the cyclol fabric into polyhedra. The cyclol fabric possesses many features which suggest that it can account for many of the properties of the globular proteins. Whereas the polypeptide chain, with its free rotation about the bonds connecting its links, has no definite shape and thus has no predisposition to give crystals, the cyclol fabric has a skeleton with a highly organized and symmetrical structure which might confer on molecules built from it a strong tendency to crystallize (item 9). The trigonal symmetry of the fabric might be the basis for the occurrence of the factor 3 in Bergmann's numbers (item 6). The unique position of the globular proteins among all products that are obtained by their denaturation and degradation could thus be interpreted as an indication that they alone are built exclusively of *B* units; on this view denaturation results from the breaking of cyclol bonds, with the production of less symmetrical and less closely integrated structures.

§ 6. CAGE STRUCTURES

The plane cyclol fabric, however, can extend indefinitely in two dimensions and so does not give discrete molecular weights or definite, chemically individual, substances. In this respect it has the same inadequacy as the polypeptide chain. Dr Wrinch was thus led to study the possibility of folding the plane cyclol fabric into space-enclosing structures.

In the skeleton of the cyclol fabric all the carbon atoms are surrounded by four other atoms in positions corresponding to the four tetrahedral valency angles. In the case of the nitrogen atoms, however, only three of these positions are occupied. It thus happens that the cyclol fabric can be folded along certain trigonally arranged lines, but only at the tetrahedral angle of 109°, by the shifting of bonds on certain nitrogen atoms to the previously unoccupied positions. The cyclol theory thus leads to the consideration of those polyhedral cage structures which can be formed by folding the fabric through angles of 109°. The necessity of preserving the valency angles presents certain difficulties at the corners where two folds meet. This further restricts the possible lines along which folding can occur.

The polyhedra that can be formed from the cyclol fabric in this way have definite sizes and shapes. Even after much study only one series of polyhedral cage structures has been discovered. These are all truncated tetrahedra, whose eight faces are all parallel to corresponding faces of a regular octahedron. Their sizes are such that they can comprise only 72, 288, or in general $72n^2$ residues, where n is any integer. For convenience these structures are represented by the symbols C_1, C_2, and C_n. Figure 2 shows a photograph of a model of the C_2 structure.

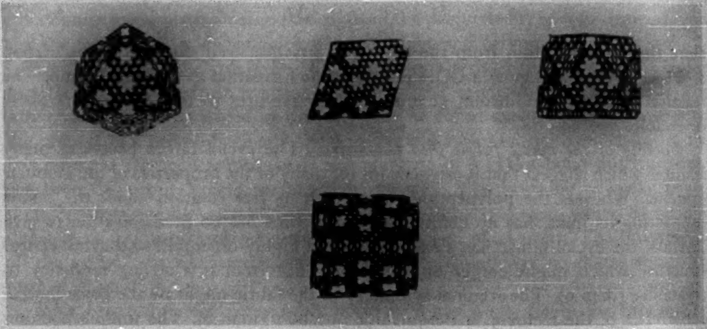


Figure 2.

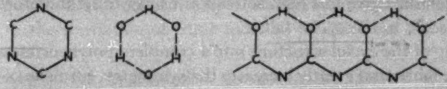
These C_n structures have no centre or plane of symmetry. If we consider only the skeleton (disregarding the positions of the side chains) they have four triad axes and three dyad axes, which are direct consequences of the triadic and dyadic symmetries of the triazine and diazine rings. With different selections and arrangements of R groups the symmetry may be decreased. Evidently, appropriate arrangements of R groups can give trigonal molecules.

These polyhedral cage structures are a mathematical consequence of the original assumptions of the cyclol theory. They show that the existence of globular proteins, having definite molecular weights, can be deduced from this theory and that no continuous variation of molecular weights of proteins is to be expected. They indicate a deep-seated symmetry in the architecture of protein molecules, which is a direct consequence of the rhythm of $C—C_\alpha—N$ in the chemical composition of proteins. Numbers such as 5 or 7 do not put in an appearance in such a picture, and numbers involving 2 and 3 occur naturally, without any *ad hoc* assumption.

These structures explain the definite molecular weights found by Svedberg, since cages comprise definite numbers of residues. The molecular-weight classes thus indicate structures having certain special numbers of residues. Different selections of amino acids cause the spread of the molecular-weight classes. The geometrical relations between the various cage structures explain the relations between the molecular-weight classes. The molecular-weight class at about 36,000, in particular, is interpreted as indicating the C_2 structure which contains 288 residues. This is in agreement with Bergmann's claim that the egg-albumin molecule contains 288 residues. The symmetries of the cyclol molecules explain the numbers $2^n.3^m$ which Bergmann has found for the numbers of individual amino acids in several proteins (item 6).

Stability of cyclol cages. Although cyclol bonds in isolation may well be unstable, yet in the cyclol fabric they may be stabilized by several different factors. The formation of a triazine ring, for example, presumably has such an effect.

In a triazine ring the three carbon and the three nitrogen atoms occupy alternate positions. The three carbons carry oxygens, and according to the original cyclol hypothesis each oxygen carried a hydrogen. The chemist, however, has found no evidence of hydroxyl groups of this kind in proteins and is thus sceptical of their existence. But there have already been many cases in which the chemist has not been able to find hydroxyls which he had reason to suppose were present in a structure. In salicylaldehyde and nitrophenol, for example, spectroscopic investigations showed that the hydroxyl band is absent. The deduction was then drawn that the hydrogen acts as an intramolecular bond between oxygens(14). This explanation fitted well with the abnormal closeness of approach (about 2·5 A.) of the oxygen atoms in these structures which in many cases had already been interpreted to indicate hydrogen bonding of atoms.



(a) bottom (b) top (c) sides

Figure 3.

Let us then treat in exactly the same way the triazine rings in the cyclol fabric, in which the oxygens also lie abnormally close together (about 2·45 A. apart). If we postulate that each hydrogen atom takes a position between two oxygens and forms a bond between them, we have a cage-like structure comprising five six-membered rings as illustrated in figure 3.

Figure 3 (*a*) shows the projection on to the plane of the fabric of the positions of the three carbon and three nitrogen atoms that form the skeleton of the triazine ring. If we consider one of the triazines in which the oxygens lie above the carbons, then the top of the cage, which is shown in figure 3 (*b*) as a projection on to the plane of the fabric, consists of a six-membered ring having three oxygens connected by three hydrogen bridges. The three sides of the cage are shown diagrammatically in figure 3 (*c*) as a developed surface formed by opening the triazine ring and spreading its sides out into a single plane perpendicular to that of the fabric.

The presence of hydrogen bonds between the oxygens of the triazine rings would not only account for the short distance between the oxygen atoms and the chemist's failure to detect OH groups in proteins, but would give a type of resonance that could increase the stability of the cyclol fabric and, by permitting the transmission of chemical effects throughout the fabric, explain the extraordinary specificity of the globular proteins.

In the cyclol fabric, because of the multiple paths by which any two atoms of the skeleton are bound together, the breaking of a single cyclol bond does not lead to a disintegration of the fabric. The broken bond may thus be re-established at a later time. It is reasonable to assume that throughout the cyclol fabric there may be a tautomeric equilibrium between *A* and *B* units by which cyclol bonds are continually breaking and reforming, although at any time only a negligible fraction of *B* units are present. This would give to the proteins a high reactivity. Denaturation which depends upon a breakdown of the fabric would thus require the simultaneous breaking of a large number of cyclol bonds and would involve a very high activation energy in accordance with the extremely high observed value of the temperature coefficient of the rate of denaturation by heat. The multiple paths of linkage which thus give this high activation energy must therefore also contribute a stabilizing factor to the cyclol fabric.

The closing of the cyclol structure into a complete polyhedron, which eliminates edges from the fabric and greatly increases the symmetry, furnishes a third stabilizing effect, which would help to overcome any inherent instability in isolated cyclol bonds. The chemical objections that have been made to the cyclol bond on the score of stability are therefore not applicable to the cyclol structure of the globular proteins.

Resonance within cyclol cages. With the hydrogen bonding between the oxygen atoms, each triazine ring becomes a resonant system which links together six neighbouring side chains. May we not perhaps regard each diazine ring in the cyclol fabric as the analogue of an electrical circuit, whose natural frequency is determined by the character of the side chains in this ring? If we have *n* resonant circuits coupled together through the triazine rings they should be characterized by n^2 frequencies, no one of which would be the same as the natural frequencies of the separate circuits. An electrical network having coupled resonant circuits with the geometrical arrangement of the diazine rings of the cyclol polyhedra should possess interesting properties that might help to explain the specificity of the globular proteins. The situation would presumably be somewhat analogous to that in the Debye theory of the specific heats of crystals, in which a crystal is considered as having a large number of modes of vibration. Thus the cyclol polyhedra may be characterized by a whole spectrum of frequencies definitely correlated with the symmetry of the structure as a whole. A single factor which disturbs the symmetry might have a profound effect upon some of the characteristic frequencies. If we imagine that these frequencies are important in the interactions between proteins or in funneling the energy to a prosthetic group (as to the haematin in haemoglobin), we have a possible reason for the important effects produced by apparently minor changes in the structure of the protein molecule.

Protein monolayers. The spreading of proteins to form monolayers which are extremely insoluble, even when compressed between barriers, proves the presence within each molecule of protein of hydrophobic groups having an effect equivalent to at least 340 CH_2 groups(6). On the other hand, the high solubility of many globular proteins in water shows that these hydrophobic groups cannot be present in the surfaces of the molecules but must be hidden away in the interior. The solubilities of the globular proteins and the insolubilities of their monolayers is definite evidence for a cage structure of the molecules of the globular proteins. The breakdown of the cage structure by the formation of the monolayers or by other forms of denaturation makes available the hydrophobic groups and so tends to render the protein insoluble.

The situation in regard to the solubilities of the proteins is closely related to the solubilities of soaps and many other substances whose molecules have long hydrocarbon chains which carry ionic groups at one end(7). The hydrophobic parts of the molecules of these substances tend to coalesce and thus form droplets or micelles having a hydrophilic surface and a hydrophobic interior. G. S. Hartley has shown that cetyl trimethyl ammonium salts form spherical micelles of nearly uniform size having a radius equal to the length of the hydrocarbon chain(15). The decrease in free energy involved in the formation of these micelles results from the coalescence of hydrophobic groups.

The coalescence of the hydrophobic groups in the interior of a cyclol cage molecule, such as that postulated for insulin, would similarly result in a large decrease in energy and so furnishes us with a fourth stabilizing factor for the cyclol polyhedra. A rough calculation indicates that at least 600 kcal. is involved in bringing together the hydrophobic groups of an insulin molecule within the cyclol cage structure.

Denaturation of globular proteins. The polyhedral cage structure corresponding to C_2 contains 288 residues. The number of ways in which 20 kinds of side chains could be arranged in these 288 available positions is of course prodigious, but is insignificant in comparison with the number of products that could be formed by the progressive breakdown of the structures by the successive breaking of cyclol bonds. Thus the globular protein, which corresponds to the cyclol cage, occupies a unique position as the only structure consisting wholly of B units. No such unique position of the globular protein is suggested by the polypeptide theory or, as far as I know, by any other theory of protein structure. It fits the observed remarkable properties of the globular proteins so well as to give strong support to the cyclol theory.

It is probable that very special conditions, characteristic as yet only of living organisms, are required to bring amino acids into the highly organized structure of the cyclol polyhedron. Denaturation corresponds to any breakdown of the structure produced by the opening of cyclol bonds. We can see why denaturation is usually an irreversible process, although in some cases of very mild denaturation, where few bonds are opened, the bonds may reform because the multiple paths of binding hold the parts in suitable position. Sufficiently severe denaturation or degradation treatments yield finally the polypeptide chains which the chemist finds.

§7. GENERAL EVIDENCE FOR THE CYCLOL THEORY

We have now compared the properties that are to be expected from the cyclol theory of the structure of the globular proteins with the first eleven of the items of our list of properties of these substances. In my opinion, the agreement has been so good as to justify the cyclol theory and to give a strong reason for believing that the theory gives the true structure of these proteins.

The question naturally arises whether other possible structures might not fit the facts equally well. The very large number of independent properties which are correlated by this theory, and the extreme simplicity and small number of postulates from which it is derived by purely logical processes, lead me to believe that for the simplest globular proteins it is improbable that any essential modification of the present cyclol cage structure can be made which will fit the facts as well. For example, it has been suggested that the cyclol bond, instead of being a covalent bond between carbon and nitrogen, might involve a hydrogen bond. Any such modification, however, would reduce the symmetry of the structure and would increase the distance between the atoms so that the C_2 cage polyhedron would have a volume far larger than the known volumes of the protein molecules.

It must be recognized that the cyclol theory in its present form cannot assign definite positions to particular amino-acid residues. It is a merit of the theory that it leads to a characteristic pattern for the skeleton of the simplest globular proteins without requiring knowledge of the side-chain arrangement. Its indication that these proteins have many features in common, as given in our list of 12 items, independently of the selection of side chains, is well supported by the facts.

X-ray evidence regarding the structure of globular proteins. The physical and chemical evidence in support of the cyclol cage structure for the molecules of globular proteins which we have discussed is of the same general character and sufficiency as that which has been found so useful by chemists in establishing the correct structures of organic molecules. The modern methods of X-ray analysis of crystals and molecular structures, when sufficiently detailed and accurate data are available, have frequently given a precise and certain determination of structure. In most cases these structures have confirmed the findings of the chemist, but sometimes they have led to modifications which have thrown new light on the nature of chemical combinations, as for example in structures involving hydrogen bonds and other resonant systems. It is probable that the final decision regarding the correct and complete structures of globular proteins will be based upon X-ray data.

§8. THE STRUCTURE OF INSULIN

Among the six proteins which have given good X-ray pictures only one, insulin(16), shows that the unit crystallographic cell contains a single molecule. Since the interpretation of the data is greatly facilitated by the knowledge that the cell contains only one molecule, the case of insulin assumes particular importance as a test of the cyclol theory. Let us examine therefore the conclusion that can be drawn from the X-ray studies of insulin.

The crystal is found to have trigonal symmetry, and therefore since there is only one molecule per cell the molecule itself must have trigonal symmetry. Such high symmetry is very unusual in large organic molecules even though these are much smaller than the insulin molecule. We have seen that the skeleton of the cyclol cage polyhedron has trigonal symmetry, so that with suitable choice of side-chain locations the molecule itself could be trigonal. The cyclol structure, in thus explaining a type of symmetry unusual for large molecules, receives confirmation from the x-ray data.

Sedimentation data obtained with the ultracentrifuge have shown that insulin molecules are approximately spherical and belong to the 36,000-molecular-weight class to which the C_2 structure with 288 residues should apply. If a mean distance 1·5 A. is taken between carbon-carbon and carbon-nitrogen atoms, the absolute size of the C_2 cage can be calculated; it is found to fit the cell given by the x-ray measurements, which it would not have been likely to do if any other than the cyclol bond had been postulated$^{(17)}$.

Subsequently Crowfoot measured the relative intensities of reflection from various planes, giving a total of 59 intensity-measurements$^{(18)}$. From these data she was not able to determine the structure, but summarized her data in five vector diagrams which gives the relative distribution of the vectors connecting all pairs of electrons in the crystal. The complete 3-dimensional vector map, of which the sections and projections given by Crowfoot furnish a rather imperfect picture, should contain all the data available from the x-ray measurements. To obtain the structure from the vector map requires some assumptions as to the nature of the centres that give the x-ray scattering. Dr Wrinch and I have assumed the smallest possible number of discretely arranged high-density or low-density regions, which we propose to call discrons, that seem capable of accounting for the main features of the Crowfoot diagrams$^{(19)}$. Our studies of the data show that the most important maxima of the Crowfoot diagrams can be interpreted as being due to six high-density discrons per molecule located at the six corners of a regular octahedron of side 29·4 A. Now the six slits or corners of the C_2 polyhedron, as shown in figure 2, also lie at the six corners of a regular octahedron of side 29·4 A. Therefore if we assume at these slits discrons of sufficiently high density, we find that the x-ray data give direct confirmation of the size and shape of the cyclol polyhedron as given by the location of its corners.

Further examination of the Crowfoot diagrams shows features that correspond to (1) a low-density discron at the centre, i.e. a hollow cage; (2) low-density discrons at the positions corresponding to the lacunae of the cyclol fabric; and (3) high-density discrons near the middles of lines connecting molecular centres (except along the z axis) where the zinc atoms, with associated polar side chains that bind the molecules together, are probably located.

The x-ray data, in my opinion, are thus not incompatible with the cyclol structure for insulin, but on the contrary give strong confirmatory evidence of its correctness.

Unfortunately, Crowfoot's data give only relative intensities, so that one cannot determine how many atoms of, say, sulphur in each of the discrons at the slits would be sufficient to account for the features of the X-ray pictures, nor at present can one analyse in detail the nature of the low-density discron at the centre. Since the average density of the electrons in the proteins is not greatly different from that of water, which must fill all intermolecular spaces, the X-ray reflections are produced by slight variations in electron-density associated with certain atoms and groups of atoms in the skeleton and side chains. The C—C_α—N skeleton of the cyclol fabric represents only 28 per cent by weight of the insulin crystal. Since the most uniform distribution of electron-density is in the skeleton, the latter probably contributes less than the side chains to the rather definite maxima and minima shown in the vector diagrams. It seems probable that the observed discrons at the slits are due mainly to sulphur atoms in side chains located in these positions. The low-intensity region at the centre is due to the fact that the distance between parallel faces of C_2 is 24 A., so that with any arrangement of side chains there is necessarily a region at the centre which they do not fill. This region may be filled with water or it may even be a void.

The system of given discrons located at the corners and centre of an octahedron of side 29·4 A., suggested by the C_2 structure, was also deduced from the published (0001) vector projection(20). A second (0001) vector projection, corrected for temperature effects (so far unpublished), was produced by Bernal at the Royal Society discussion on proteins on 17 November 1938. This projection permits, although it does not require, the displacement of the *A*, *B*, *C* maxima to degenerate positions which lie on a single equitriangular mesh system, with a side of about 10 A. common to the whole crystal. Adopting this placing of the peaks, Bernal has suggested(21), as the structure of insulin, a set of 18 discrons whose projections on the (0001) plane lie on this equitriangular mesh system. These discrons are consistent with this particular reading of the (0001) vector diagram, but they must also be considered in relation to the three-dimensional vector map. Of these 18 discrons, six of course are located at the corners of the C_2 octahedron as before, the six which lie outside the C_2 structures are not confirmed by the $z=0$ vector section and so must be discarded, and the remaining six (whose existence in view of the $z=0$ vector section is by no means certain) can without difficulty be interpreted in terms of the C_2 structures. There is therefore no inconsistency between the C_2 structure and the new (0001) vector diagram, whether or not the *A*, *B*, *C* peaks lie in the special positions assumed by Bernal to be the correct one(21).

In view of the fundamental importance of the structure of insulin more extensive data are urgently required. It must be recognized that 59 relative intensities are not enough to establish the structure of insulin. Thus it is not possible from relative intensities to make a quantitative determination of the intensities of the postulated discron system necessary to account for the vector diagrams. To discover, for instance, how many sulphur atoms must be placed in each of the octahedrally arranged discrons, it is necessary to have absolute intensities. Further, these data relate to a dry crystal. In view of the remarkably detailed X-ray photographs of wet haemoglobin(4), it is probable that a wet insulin crystal would yield a very much more extensive set of data. Again, the diagrams relate to a zinc-insulin crystal only.

But insulin is known to crystallize also with cadmium and with cobalt atoms in stoichrometric proportions. A study of wet and dry insulin crystals, in which cadmium or cobalt or even, if possible, mercury took the place of zinc, would be of great value.

§ 9. CONCLUSION

We may sum up the present position with regard to the structure of proteins as follows. A vast amount of data relating to protein structure have been collected by workers in a dozen different fields. No reasonable doubt remains as to the chemical composition of proteins. The original idea of native proteins as long chain polymers of amino-acid residues, while consistent with the facts relating to the chemical composition of proteins in general, was not a necessary deduction from these facts. Moreover it is incompatible with the facts of protein crystallography, both classical and modern, with the phenomena of denaturation, with Svedberg's results which show that the native proteins have definite molecular weights, and with the high specificity of proteins discovered in studies in immunochemistry and enzyme chemistry. All these facts seem to demand a highly organized structure for the native proteins, and the assumption that the residues function as two-armed units leading to long-chain structures must be discarded. The cyclol hypothesis introduced the simple assumption that the residues function as four-armed units, and its development during the last few years has shown that this single postulate leads by straight mathematical deductions to the idea of a characteristic protein fabric which in itself explains the striking uniformities of skeleton and configuration of all the amino-acid molecules obtained by the degradation of proteins. The geometry of the cyclol fabric is such that it can fold round polyhedrally to form closed cage-like structures. These cage molecules explain in one simple scheme the existence of megamolecules of definite molecular weights capable of highly specific reactions, of crystallizing, and of forming monolayers of very great insolubility. The agreement between the properties of the globular proteins and the cyclol structures proposed for them is indeed so striking that it gives an adequate justification for the cyclol theory, especially in view of the fact that this great variety of independent facts are on this theory seen to be logical consequences of one simple postulate.

REFERENCES

(1*a*) Bergmann and Niemann. *J. Biol. Chem.* **115**, 77 (1936).
(1*b*) Bergmann and Niemann. *J. Biol. Chem.* **118**, 301 (1937).
(2) Schimper. *Z. Kristallogr.* **5**, 131 (1881).
(3) Bernal. *Reunion International de Physique, Chimie et Biologie*, Paris, p. 408 (October 1937).
(4) Bernal *et al. Nature, Lond.*, **141**, 521 (1938).
(5) Langmuir and Schaefer. *J. Amer. Chem. Soc.* **60**, 1351 (1938).
(6) Langmuir. *Cold Spr. Harb. Symp. Quan. Biol.* **6**, 171 (1938).
(7) Langmuir. Pilgrim Trust Lecture, Royal Society, *Proc. Roy. Soc.* A, **170**, 1 (1939).
(8) Langmuir. *Chem. Rev.* **6**, 451 (1929) (see especially p. 466).
(9) Wrinch. *Nature, Lond.*, **134**, 978 (1934); *Protoplasma*, **25**, 550 (1936).
(10*a*) Wrinch. *Nature, Lond.*, **137**, 411 (1936) *et seq.*

(10*b*) Wrinch. *Proc. Roy. Soc.* A, **160**, 59 (1937).
(11) Astbury. *J. Text. Inst., Manchr*, **27**, 283 (1936).
(12) Langmuir and Wrinch. *Nature, Lond.*, **143**, 49, fig. 2 (1939).
(13) Wrinch. *Nature, Lond.*, **139**, 972 (1937); *Proc. Roy. Soc.* A, **161**, 505 (1937).
(14) Hilbert, Wulf, Hendricks and Liddel. *Nature, Lond.*, **135**, 147 (1935).
Errera and Mollett. *C.R. Acad. Sci., Paris*, **200**, 814 (1935).
(15) Hartley and Rummicles. *Proc. Roy. Soc.* A, **168**, 401 (1938).
Hartley. *J. Chem. Soc.* p. 1968 (1938).
(16) Crowfoot. *Nature, Lond.*, **135**, 591 (1935).
(17) Wrinch. *Science*, **85**, 568 (1937); *Trans. Faraday Soc.* **35**, 1368 (1937).
(18) Crowfoot. *Proc. Roy. Soc.* A, **164**, 580 (1938).
(19) Wrinch. *J. Amer. Chem. Soc.* **60**, 2005 (1938).
Wrinch and Langmuir. *J. Amer. Chem. Soc.* **60**, 2247 (1938).
(20) Wrinch. *Nature, Lond.*, **142**, 956 (1938).
(21) Bernal. *Nature, Lond.*, **143**, 74 (1939).
(22) Langmuir and Wrinch. In the Press.

DISCUSSION

Dr D. Wrinch: I do not want to-night to go into any details regarding the theoretical aspects of protein structure, but rather to thank Dr Langmuir for the new light which he has thrown on the subject. I would also point out how critical it is for the future progress of our knowledge of protein structure—upon which of course the future of medicine also depends—that additional data should be obtained. It is perhaps hardly realized by workers in physics that we do not yet know the complete composition of any single globular protein. Data regarding the chemical composition of insulin are still very incomplete and make it impossible at present to locate the side chains. Such chemical data are urgently needed and can be obtained by the application of the established methods of chemical analysis. Dr Langmuir has stressed the great importance of applying to the proteins all the techniques of physics which are appropriate. Great progress in the isolation and crystallization of proteins now provides for physical investigations an almost unlimited wealth of material.

A NOTE ON THE STRUCTURE OF INSULIN

By IRVING LANGMUIR and DOROTHY WRINCH

Received 13 *December* 1938. *Read* 12 *May* 1939

ABSTRACT. The modified (0001) vector projection for an insulin crystal is examined in relation to the C_2 structure proposed for the insulin molecule. It is found that this projection confirms the C_2 structure represented by the discron system $3s+3s'-o$ provided it is adequately loaded at its slits. Bernal's suggestion of an 18-discron set structure for the insulin molecule is shown to be untenable, since it is inconsistent with the $z=0$ vector section.

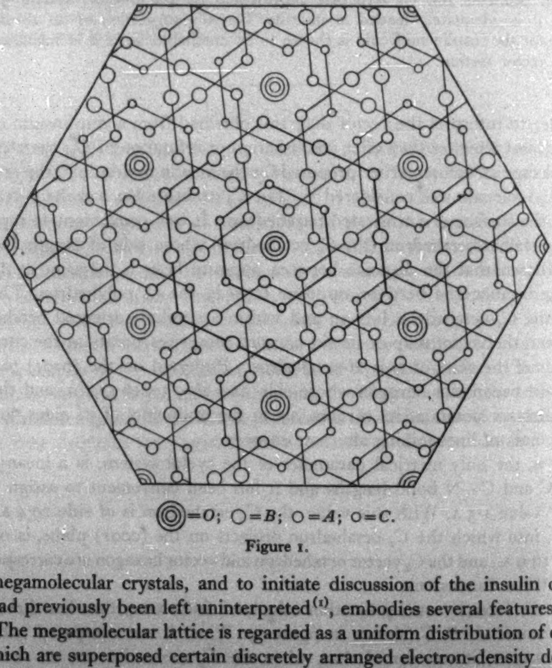
In order to interpret the vector diagrams obtained from a zinc-insulin crystal(1), and to test whether they offer confirmation or refutation of C_2, the 288-residue cyclol cage structure earlier proposed for the insulin molecule(2), the vector map of this C_2 molecule was considered(3). The C_2 structure has a skeleton framework lying on the surface of a truncated tetrahedron. It was convenient to replace this by an equivalent octahedron (the C_2 octahedron) whose side of length $l=8\sqrt{6}.a$ was so chosen that the distance between opposite faces is the same as the corresponding distance $16a$ between opposite faces in the C_2 polyhedron. The vector map of the C_2 octahedron lies on and within a similarly oriented octahedron of side $2l$, and the vector map of its six corners comprises points at the corners and midpoints of the sides of this $2l$-octahedron. Projected on the (0001) plane, this octahedron becomes a hexagon whose side $2h=2l/\sqrt{3}=16\sqrt{2}.a$, and the points project into six points at its corners, six at the midpoints of its sides, and six at the midpoints of lines joining alternate corners.

Now a, the only metrical parameter of the cyclol system, is a mean between the C—C and C—N bond lengths and it has been convenient to assign to it the arbitrary value 1·5 A. With this value, the C_2 octahedron is of side 29·4 A., the C_2 hexagon, into which the C_2 octahedron projects on the (0001) plane, is of side h, equal to 16·9 A., and the C_2 vector octahedron and vector hexagon are correspondingly of twice these dimensions.

Now the (0001) vector projection calculated from X-ray data has three hexadic sets of maxima A, B, C (shown in figure 1). When it was compared with the C_2 vector hexagon, it was found that, when the C_2 molecules are set skew in the lattice at a certain angle α of about 6°, the A, B, C points are the vector map of $3s+3s'-o$, namely a triad of high-density regions of intensity s at the three upper corners of the C_2 octahedron, a second triad of intensity s' at the three lower corners, and a low-density region $-o$ at its centre, the negative region at the centre accounting for the low intensity of the C points(4). We conclude that this vector projection is consistent with the proposed C_2 structure (which necessarily has a low-density region at its centre) provided that there are adequate concentrations of atoms, belonging presumably to special side chains at its six corners.

The vector map of a crystal is a three-dimensional inventory of all the information obtained by means of x-ray observations. This was emphasized in another investigation(5) in which the location in three-dimensional vector space of interactions of these, and certain other regions of high and low electron-density suggested by the structure of C_2, were checked from the various available sections of the vector map.

This interpretation of the experimental diagrams, put forward at an early stage of the work as a demonstration of the application of the method of discrete point set to megamolecular crystals, and to initiate discussion of the insulin diagrams which had previously been left uninterpreted(1), embodies several features.



◎=O; ○=B; ○=A; ○=C.

Figure 1.

(1) The megamolecular lattice is regarded as a uniform distribution of electrons upon which are superposed certain discretely arranged electron-density deviations or *discrons*. A discron is an isolated region in which the electron-density is high (in which case the discron is positive and can represent a single electron, an atom, or a group of atoms) or low (in which case the discron is negative and can represent a region sparsely populated with electrons or even a hole or void).

(2) The vector diagrams were shown to confirm the suggestion(2) that the insulin molecule has a structure of highly symmetric type.

(3) The skewness of the A, B, C points in the (0001) plane was interpreted to mean that the symmetric molecules are set skew in the crystal at a small angle α of about 6°.

The preliminary successes in checking the predicted structure by the experimental diagrams led to a development on geometrical lines of the general point of view of crystal megalattices and their vector maps as sets of discrons in atomic space and vector space respectively(6). The crucial question is then, evidently, the interrelation of the two types of discron sets(7).

Assuming that the A, B, C points of the (0001) projection shown in figure 1 are due to interactions of discrons, we deduce from it two and only two point sets in atomic space(8). One of these is necessarily discarded, since the relations between the intensities in the vector projection ($B > A > C$) do not allow any discrons at these points to tally satisfactorily with the given intensities in vector space. It was therefore deduced that the A, B, C points of the (0001) vector projection implies the discron set $(3s + 3s' - o)$ corresponding to the projection of the centre and octahedral points of the C_2 molecule (figure 2). By considering the various sections, it was seen that these points actually lie at heights appropriate for the C_2 octahedron.

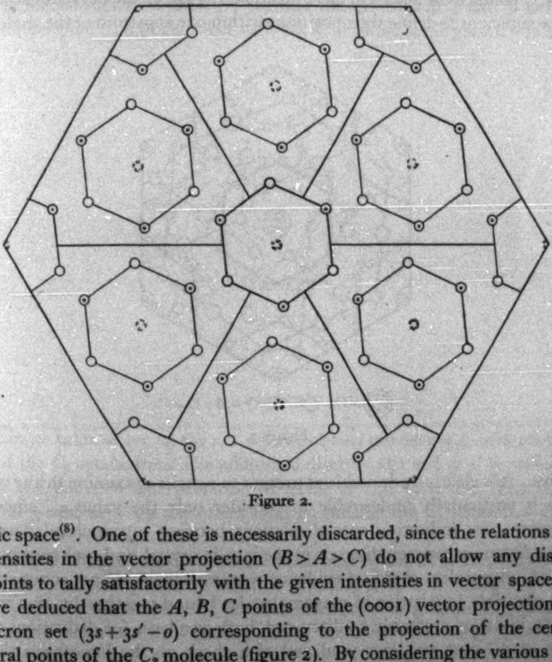
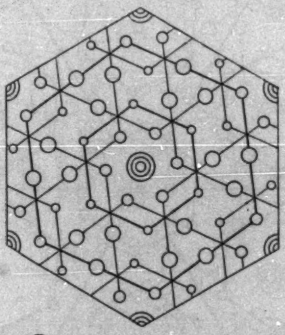


Figure 2.

Subsequently attention was directed(9) to the strong intensity of the reflection from the $(8, \bar{7}, \bar{1}, 0)$ plane which cuts the (0001) plane at the angle $\tan^{-1}(1/5\sqrt{3})$ or 6° 36′ to the axes of the crystal. But this datum (in common with all the other data obtained) is already taken account of in the vector diagrams. It, in particular, is manifested in the (0001) projection in the approximate alignment of the A, B, C

points at a small angle α to cell boundaries and thus gives no additional information. Further, a new (0001) projection, corrected for temperature effects, was produced(9) and the three sets of points A, B, C were placed at points on a single equitriangular mesh system common to the whole crystal, set in the lattice at the angle α_0 equal to $\tan^{-1}(1/5\sqrt{3})$, and having its side d equal to $b \sin \alpha_0 / \sin \pi/3$ or $b/\sqrt{57}$, where b, the side of the hexagonal cell, is 74·8 A. and d is consequently 9·92 A. This mesh may be denoted by (d, α_0). From the original report it was not possible to define α more closely than as "about 6°", since the numbers obtained by calculation from the X-ray observations were not given and only sketches of contour lines were shown. While it may be the case that, as has been assumed(9), the numbers allow the A, B, C points to be placed in these positions (figure 3), the degree of accuracy can hardly be sufficient to define their positions within one angstrom or the angle α within one degree. It is clearly undesirable at present to restrict discussion to any one value of α. It is particularly undesirable to consider only the value α_0, since this is a degenerate case, indicating special relations between intramolecular and intermolecular distances between discrons in the insulin crystal, and such relations, in the absence of independent chemical evidence requiring them, do not commend themselves as plausible in a lattice of megamolecules bound together by zinc ions. The case α_0 can however profitably be discussed, both as a basis for deductions as to the structure of the zinc-insulin crystal and as a check on the hypothesis that C_2 is the structure of the insulin molecule. We notice first that when the points A, B, C are placed on the points of the mesh system (d, α_0), the (0001) projection permits the postulation of discrons at all points of this same mesh system in atomic space. Since the zinc-insulin crystal is trigonal, all discrons except those on the trigonal axis z fall into triads, the discrons of one triad having the same intensity and the same z co-ordinate. Thus the discrons under consideration comprise, in addition



◎=O; ○=B; ○=A; ○=C.

Figure 3.

to discrons at 0, discron triads at certain heights at 1, at 1′, at 2, at 2′, at 3, and at 3′, as shown in figure 4.

In presenting the new projection, Bernal adopted the features (1), (2) and (3) mentioned above and drew attention to the eighteen points 1, 1′, 2, 2′, 3, 3′, suggesting that discrons at these points are permitted by the x-ray evidence. But while the new placing of A, B, C in the (0001) projection permits discrons to be postulated at all these points, it does not require this and consequently the existence of each discron must be confirmed in turn by reference to the three-dimensional vector map.

Comparing this system of discrons with the system originally postulated in accordance with the hypothesis that the insulin molecule has the structure C_2 adequately loaded at its six slits, we see that the discrons at 2 and 2′ correspond to the discrons s and s', a being taken as 1·515 A. or (if for some reason it is desirable to maintain its value at 1·5 A.) the s, s' discrons being moved 0·5 A. inwards from the corners of the C_2 octahedron. The additional discrons at 1 and at 1′ lie within the C_2 structure and allow of interpretation in terms of this structure. The remaining discrons at 3 and at 3′ on the other hand lie outside the C_2 structures and there appears to be no reason to expect them.

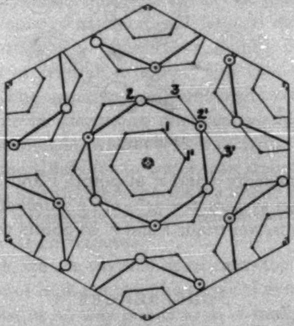


Figure 4.

To discover which of the discrons are confirmed by the three-dimensional vector map, we notice that any discron triad of intensity i with the same z co-ordinate necessarily gives an interaction of positive intensity i^2 at a hexad of points on the $z=0$ plane in vector space. Figure 5 shows the $z=0$ vector diagram. On it are marked the points 1^2, 2^2, 3^2 at which the domestic interactions of triads at 1 and at 1′, at 2 and at 2′, at 3 and at 3′, respectively should occur. The points 2^2 lie in high-intensity regions, in accordance with the postulate of s, s' discrons at the corners of the C_2 octahedron. The points 3^2 on the other hand lie in low-intensity regions, and it is therefore necessary to reject the hypothesis that there are discrons at the points 3 and 3′. The points 1^2 lie in regions of moderate intensity, leaving it for the moment an open question whether discrons at 1 and 1′ exist or not.

This preliminary investigation of the degenerate case α_0 thus yields the conclusion that the only discron sets for the insulin crystal permitted by the points A, B, C consists of the discron system $3s+3s'-o$, with the possible addition of discron triads of lower intensity r and r' at points 1 and 1′. Thus the C_2 structure represented by the discron system $3s+3s'-o$, i.e. *adequately loaded at its slits* (with the possible addition of discron triads r and r' at points within the C_2 structure), is confirmed.

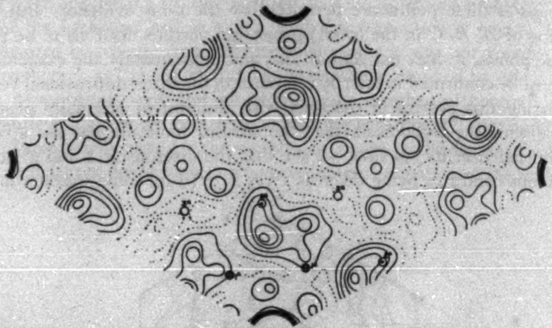


Figure 5.

REFERENCES

(1) Crowfoot. *Proc. Roy. Soc.* A, **164**, 580 (1938).
(2) Wrinch. *Science*, **85**, 566 (1937); *Trans. Faraday Soc.* **33**, 1368 (1937).
(3) Wrinch. *J. Amer. Chem. Soc.* **60**, 2005 (1938).
(4) Langmuir and Wrinch. *Nature, Lond.*, **142**, 581 (1938).
(5) Wrinch and Langmuir. *J. Amer. Chem. Soc.* **60**, 2247 (1938).
(6) Wrinch. *Phil. Mag.* **27**, 98 and 490 (1939).
(7) Neville. *Nature, Lond.*, **142**, 994 (1938).
(8) Wrinch. *Nature, Lond.*, **142**, 956 (1938).
(9) Bernal, Royal Society discussion on proteins 17 November 1938.

DISCUSSION

Prof. J. D. **Bernal**. I very much regret that I am unable to accept the major conclusions to which Dr Langmuir's address leads. Dr Langmuir maintains that apart from the x-ray evidence all the physicochemical evidence points to the cyclol model and in fact provides practical proof of it. In my opinion, on the contrary, of the eleven points which he cites as proof of the cyclol hypothesis, only one lends some doubtful support to the hypothesis, five have been assumed in framing the hypothesis and cannot be claimed as confirmation of it, one is irrelevant, and the remainder, although they can be explained by the cyclol hypothesis, can equally well be explained in a number of other ways.

The cyclol hypothesis assumes the peptide residue (1), the existence of amino-acids (2), and the tetrahedral configuration of the carbon atoms (4). Dr Langmuir's third point, the laevo configuration of the acids, though interesting, does not lead to any results which are used in the cyclol hypothesis, as the actual position of the side chains is left indeterminate. The idea of a spherical molecule (6) of definite size (5) and packing of crystalline aggregates (9) were all points which have been known before the setting up of the cyclol cage structure and in fact led to it. Only the fact that one of the cyclols contains 288 residues, which may be but by no means necessarily must be the number of residues in such a protein crystal as insulin, can be cited as any indication that the cyclol hypothesis predicts observed properties of protein molecules. The crystallinity and specificity of proteins (9 and 10) are mere consequences of the existence of discrete and identical molecules, while the fact of denaturation and the formation of monolayers simply indicates that the protein molecule has some form of regular structure, and cannot be used to justify any particular regular form. The argument of monolayers resembles in fact that condemned by Dr Langmuir in discussing chemical analysis, because the original structure of the protein molecule, whatever it was, is destroyed when once it is in a monolayer. Point 10, the solubility in water, can only be explained by an additional assumption, that all hydrophobic groups are inside and all hydrophilic groups outside, and as the cyclol hypothesis offers no reason for this assumption it cannot be claimed as a proof for its correctness.

Taken in all, therefore, we can say that apart from the assumptions of the cyclol ring involving a structure of high energy, the physical evidence, though compatible with the cyclol hypothesis, in no way points to it. I have been connected with the hypothesis since its inception, and as long as no other arguments were put forward I was content to consider it as a possible solution which had neither been proved or disproved; and as chemical evidence for its plausibility seems to be lacking the balance seems definitely to turn against it.

With the X-ray evidence, however, the position is much clearer. As Dr Langmuir said, the reality of the structure must stand or fall by the X-ray evidence alone, and in my opinion the evidence already presented is sufficient to require its definite rejection. The only convincing evidence comes from the structure of insulin. Now the formation of the cyclol cage is particularly rigid, and has a higher symmetry than is required for the structure of insulin. It is in fact much too nearly spherical to fit the definitely compressed structure of the insulin cell. The two cyclol cages along the axis would only be 7·5 A. between their faces, and this is definitely a low value to allow for any reasonable length of side chain. On the other hand, the lateral distance between the side faces is too long. The shape of the cyclols cannot be claimed to provide a particularly good fit in the case of insulin. But even more critical is the evidence of the internal structure. There must needs be considerable ambiguity in the deduction of a structure from the X-ray data, but if a hypothetical structure is put forward it is a straightforward but laborious process to calculate the X-ray patterns it should give and compare them with those observed. If they do not fit, as I maintain they do not in this case, the structure must be rejected. In the first place, as Mr Riley has shown, the vector projection of the cyclol cage when every carbon and nitrogen atom is taken into account is totally different from that observed. It contains no marked maxima, and the general positions of high and low intensity are almost completely reversed from those necessary to give agreement with the picture. The high-density *S* points assumed by Dr Langmuir and Dr Wrinch actually correspond to low-density areas in the projection. Anyone accepting the cyclol theory must therefore come to the paradoxical conclusion that its characteristic skeleton not only does not appear in the x-ray pattern, but has to be compensated by an arbitrary and improbable arrangement of scattering matter in the side chains. Whatever this arrangement may be it is demonstrably not that chosen by Dr Langmuir and Dr Wrinch. They have selected certain vectors between their assumed scattering centres, but ignored a large number of others of equal or greater intensity. The result of making a projection on the basal plane of *all* the vectors derived from the strongest scattering centres, positive and negative, which they postulate is a pattern which bears no recognizable resemblance to that observed. If any success is to be obtained along these lines it would mean such adjustment of scattering centres or discons as hardly requires and certainly cannot confirm the cyclol hypothesis.

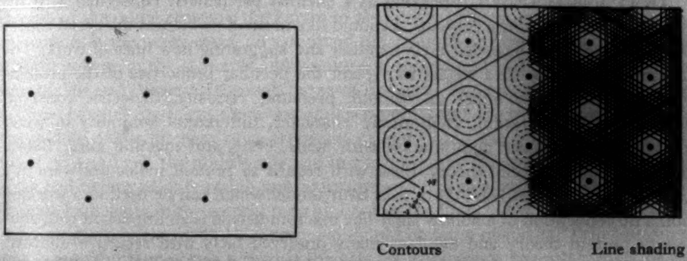
Dr Langmuir mentioned the group of arrangements which I postulated in my communication to the discussion at the Royal Society. The sum of these he points out give results incompatible with the Patterson sections. This I fully admit, but I think in connection with them there is a point to be borne in mind, that all these photographs are of dry crystals and in these cases the molecules must be considered to be slightly and irregularly tilted in relation to each other, and therefore all reflections due to scattering of points distant from each other in the molecule will tend to be blurred, while those from points close together will not be so affected. For this reason I should consider that the attribution of strong peaks in the Patterson diagram to the interference of waves from distant points in the molecule is very doubtful. I think at this stage it is premature to attempt any analysis of the molecule. We must wait for further experimental work, and I am in hearty agreement with Dr Langmuir and Dr Wrinch as to the need for this.

Dr F. C. FRANK. As Prof. Neville says,* this work has done a real service in making more widely known the use and significance of Patterson-Harker vectorial distribution plots. It is true that the P.-H. distribution does not allow of a unique solution giving the distribution in physical space. The same is identically true of the original x-ray data themselves, which contain neither more nor less information. In simple cases we obtain a unique solution by using our additional knowledge that matter is composed of atoms. The observed P.-H. distribution can sometimes be reasonably approximated to as a point set. I am informed that when this is the case it is possible to determine systematically what point-set structure or structures, if any, in physical space can correspond to it. If a unique solution is found containing as many points as atoms, these points are undoubtedly atomic sites. A point is obviously a good approximate representation of an atom: on the other hand, it is not self-evidently a good representation of the region of greatest density among a thousand atoms, so that, contrary to the conclusion of Neville's letter, for protein crystal analysis we are nearly back where we started. But not quite; if the point-set approximation to the P.-H. distribution was good enough we have derived a unique point-set structure, which *may* represent directly the distribution of electron density in physical space; but so may an infinite number of structures derived from it according to certain strict rules, namely by arbitrary changes of phase of its Fourier components, regard being paid to symmetry restrictions.

---

* *Nature* **142**, 995 (1938).

It is instructive to consider the nature of the structures which may be equivalent to the unique point set. This is illustrated in a simple two-dimensional case (which may be a projection from three dimensions) in the accompanying sketches. If the unique solution is found to be a hexagonal array of points (supposed to stand on a background of uniform positive density) as in figure 1, figure 2 is an example of structures which have identical P.-H. maps. It has been derived from (1) by a change of phase (π in this case) of 3×1 Fourier components of the infinite series representing the density distribution (1). The points stand with unaltered weights on a background which now contains periodic structure. Any periodic structure may be present, in addition to the points of (1), which can be built up out of its Fourier components changed in phase by arbitrary angles ϕ_i and multiplied in amplitude by $-2\cos\phi_i$.[*] That is to say, a periodic structure may be present, and remain invisible, provided it stands in a certain harmonic relationship to the point structure. The Fourier equivalents of a point set are in general networks containing lacunae, some of which are filled in by the points. Since there are possibilities of such periodic structure in the Wrinch model, or any other likely model of proteins, these considerations may be directly relevant.



Figures 1 and 2. Simple example of two electron-density distributions which have the same Patterson diagram. Figure 2 has been represented half in contour, half in line shading which shows separate Fourier components. A uniform positive background should be added to both figures.

[*] Because $\cos(\theta-\alpha)-\cos\theta = -2\cos\frac{1}{2}(\pi+\alpha)\cos\{\theta-\frac{1}{2}(\pi+\alpha)\}$.

Prof. Bernal counts the peculiar packing in the insulin crystal among his several serious objections to the Wrinch-Langmuir structure. This one, however, is not inconsistent with the premises of Dr Wrinch. The packing in the proposed structure may be described as follows: vertical columns of close-packed similarly oriented molecules are disposed in parallel hexagonal packing in such a way that projecting octahedron corners come close together. These corners are actually slits, and the slits meet crossed, with ends adjacent because of the 6° twist. Dr Wrinch has already proposed slit-to-slit binding to account for Svedberg's reversible polymerization. If we consider slit-to-slit binding alone, each molecule is connected to six others on levels alternately above and below it, and the crystal consists of two independent interpenetrating rhombohedral lattices. Each of these taken separately represents grouping similar in principle to those suggested by Wrinch* for Svedberg's 2, 4, 8, ... polymers.

Dr D. JORDAN LLOYD. Speaking as a chemist particularly concerned with the physical chemistry of the proteins, I should like to say how helpful the cyclol theory has been in pulling recorded facts together and suggesting new lines of work. The physical chemist needs a model to explain the peculiar properties of the proteins with regard to their effect on osmotic pressure, viscosity, dielectric constant, electrophoretic migration, zwitterionic character, differential solubility in water and alcohol, also their combination with acids, bases and metallic ions. Before the advent of the cyclol theory, ideas with regard to protein molecules were in a state of chaos; now at least there is a clear model which can be used as a working tool. The conception of a hollow shell-like macromolecule is an important contribution to protein theory and explains many observed facts with regard to density. As far as I know, the conception of a molecule formed like a hollow shell has not been advanced in any other theory. We know that the protein molecule cannot exist as an immensely long polypeptide chain wriggling and twisting about in the solution and having innumerable forms all simultaneously, since work with the ultracentrifuge has shown that in most cases all the molecules of the particular protein under investigation have the same form at the same time. This form must be rigid enough to be stable in solution but frail enough to be torn open by contact with a water-air interface.

Protein chemists have not in general the necessary mathematical equipment to take sides in any battle of mathematicians over the interpretation of vector maps and like problems. Possibly the three dimensions of the cyclol molecule may be modified in time as new knowledge accumulates. For the moment the cyclol theory is available as a working tool for protein chemists. The opponents of the cyclol theory have not up to the moment offered us any alternative. The probabilities suggest that if and when they do, it will certainly be a close relation though it may not be identical.

Prof. E. H. NEVILLE. As a mathematician I am always surprised at the lightheartedness with which my scientific friends perform extravagant extrapolations and similar illogical feats. The nature of the vector map derived from a small number of point centres is not very obscure and the reconstruction of the point configuration from such a map is very easy. But it is dangerous to assume in dealing with a megamolecule that the problem is not altered qualitatively as well as quantitatively when we have to consider not a complete map, but one which presents only the dominant features of a map that is undiscoverably complicated in detail. The assumption has been made, tacitly rather than expressly, that the dominant features of the insulin map are due to heavy concentrations in the molecule. It is true that heavy concentrations in a molecule will provide dominant features in a rough map, but the converse, as I can illustrate in a moment, is not true. Imagine a structure composed of two parallel rows of uniformly spaced points, with six points in each row, facing each other at a little distance apart. The vector map consists of three parallel rows of uniformly spaced points, with eleven points in each row; the intensities in the middle row are 2, 4, 6, 8, 10, 12, 10, 8, 6, 4, 2, and the intensities in each of the outer rows are the halves of these; the greatest intensities, 12 in one row and 6 in the other rows, do not stand out from their neighbours. Now make a slight change in the structure, displacing the second and fifth points in each row towards the end points, all through the same distance. The vector map again consists of three parallel rows, but there are nineteen points in each row; the intensities of the middle points of these rows are again 12 and 6, but the other intensities have fallen, to 4 and 2 in the middle row and to 2 and 1 in the other rows. It follows that if we had before us only rough approximations to our two maps, we should detect definite high spots in the second map and not in the first, but if we inferred atomic concentrations in the second structure we should be quite mistaken. Two congruent sets face each other across a gap; if the sets are sufficiently irregular, most of the cross vectors fade into the background, and then, because one set is an exact reproduction of the other, one cross vector stands out clearly. Here then is one way in which a vector map may show a dominant feature which is not to be assigned to concentrations in the point structure. I am not prepared to say that there are not other ways. The warning on which I insist is that all assertions or claims as to what X-ray diagrams do or do not prove about megamolecules must remain unjustified or unestablished until we have approached the problem with an open mind from the other end, intent on discovering all the causes from which dominant features can arise in an elaborate map instead of assuming that they must have arisen in some perfectly obvious fashion.

---

* *Phil. Mag.* vii **26**, 313 (1938).

Dr J. M. Robertson. I have not previously been associated with the cyclol theory of protein structure in any way, and I have no views for or against the theory other than those which I have formed to-night. Any such theory may be regarded as a useful starting point in an attempt to unravel the structure of insulin, something to be tested and accepted or rejected as the experimental facts allow. I feel sure, however, that the X-ray intensity data, on which the insulin vector maps are based, are so meagre and unavoidably incomplete that no reliable conclusion can yet be reached as to the details of this excessively complicated structure. The structures of many organic molecules, containing 20 or more atoms, have been completely worked out by the x-ray method. The 60 odd parameters defining these structures have been determined from, say, 200 independent relations between them, representing measured intensities. But if we have, instead, 200 unknown quantities and perhaps only 10 equations connecting them, then very little can be said. No unique solution is possible. The insulin situation is much worse than this. There are many thousands of atoms in the molecule and only 59 *relative* intensity measurements to work on. As far as these measurements go the structure is effectively a continuous distribution of scattering matter, for which we know that there is no unique solution. Attempts to represent the molecule by a small number of discrete scattering points do not seem to me to afford even reasonable approximations, because we cannot expect any very large gaps devoid of scattering matter between the atoms.

These destructive remarks may not be helpful, and I would like to suggest that a better interpretation of the insulin vector maps might be possible if the intensity measurements were made absolute instead of relative, for this would fix the vertical scale. Further use might also be made of the few heavy atoms in the structure, especially if these could be replaced by still heavier ones like mercury atoms. They should have the effect of swamping the x-ray reflections, and, roughly speaking, would convert unknown phase differences into differences of amplitude, which could be measured. An actual density distribution might then be mapped out directly in real space.

Dr I. LANGMUIR (in reply). I think the differences of opinion shown in the discussion are very helpful. They emphasize, as I emphasized, the need for better data and for more extensive data. I would specially stress the importance of inserting in protein crystals various heavy ions, as has been specifically proposed already for insulin.

I would also wish, in reply to Dr Bernal's suggestion that the spherical molecules C_2 were assumed in the cyclol theory, to point out that the polyhedral cage structures of the cyclol theory are a straight mathematical deduction from the cyclol postulate, the particular symmetries of the C_n cyclol cages being derived from the valency angles of carbon and nitrogen and the laevo character of the amino acids in proteins. These features of the theory are not introduced *ad hoc*. I would also explain to Dr Bernal that the work on protein monolayers indicates that the protein in solution has some radically different structure, in which very considerable numbers of hydrophobic groups can be hidden, and this provides independent confirmation of the cage structure, deduced in the cyclol theory from the geometry of polypeptide chains. After listening to this discussion my impression is still that this contest will last a relatively short time. As soon as better data are available, we shall know exactly where we stand.

THE STRUCTURE OF SILVER FILMS

By O. GOCHE and H. WILMAN
The Imperial College, London, and the University, Brussels

Received 27 *January* 1939. *Read* 12 *May* 1939

ABSTRACT. An electron-diffraction study of the structure of silver films condensed in vacuo on rocksalt cleavage faces at about 200° C., and of the changes produced by heating these films in vacuo, has shown that the reflection and transmission patterns are in general agreement with Menzer's view of the twinned structure of such films. The films examined yielded patterns of a type indicating strong {111} extensions of the reciprocal-lattice intensity regions, but without marked development of [001] extensions. No conclusive evidence has been found of development of octahedral boundary faces of the films in addition to the twinning which occurs on these planes.

The positions of the diffraction spots and lines in patterns from films which had become distorted during the dissolution of the substrate and the mounting of the film showed that parts of the film had become considerably curved about axes parallel to the cube axes and cube-face diagonal directions in the film plane. These types of curvature are in agreement with the form of the substrate surface and the twinned structure of the films. General equations are derived for use in considering the geometrical features of rotation patterns yielded by cubic crystals.

Heating in vacuo at about 500° C. quickly converts the silver films into normal relatively perfect single crystals in the same orientation as that part of the initial film whose lattice was parallel to that of the rocksalt substrate. Such an extensive atomic rearrangement testifies to a high mobility of the silver atoms at temperatures much below the melting point; this has been inferred by previous investigators, but in this case is shown to occur within the lattice and not merely on its surface.

---

§ 1. INTRODUCTION

The experiments of Lassen⁽¹⁾, Lassen and Brück⁽²⁾, Kirchner and Lassen⁽³⁾ and Brück⁽⁴⁾ appeared to indicate that silver films prepared by condensation in vacuo on heated rocksalt cleavage faces at temperatures above about 150° C. have practically a single-crystal structure, whose crystallographic axes are parallel to those of the rocksalt substrate. The electron-diffraction patterns from the thinner films (less than about 500 A. thick) contain certain abnormal streaks and spots, corresponding to diffractions with non-integral or irrational indices, which Kirchner and Lassen first attributed to a cross-grating type of diffraction by the octahedral or {111} planes of the small crystals, due to their being the planes with the greatest atom-density in the lattice. Thicker films (about 1000 A. thick) yielded only normal diffraction spots and Kikuchi lines. The abnormal spots correspond to intersections of the sphere of reflection with the diagonals of the body-centred identity cells of the reciprocal lattice, and Brück pointed out that they were strengthened wherever the intersections coincided with points dividing the cell diagonals into six equal parts as shown in figure 6.

Laue(5) has shown theoretically that even without postulating any cross-grating nature of the diffraction, such abnormal diffractions are to be expected in patterns from any kind of crystal if this be bounded by definite crystallographic planes; and that these irrational diffraction spots should appear at positions corresponding to the intersection of the sphere of reflection with spurs (*Stacheln*) extending from the reciprocal lattice points in directions normal to the boundary planes of the crystal. In the present case, as Kirchner and Lassen showed, the anomalous spots correspond to intersections of the sphere of reflection with the diagonals of the body-centred cubic cells of the silver reciprocal lattice, which are normal to the octahedral planes; hence Laue concluded that the surface of the silver film was not flat but consisted of projections bounded by octahedral faces. Laue and Riewe(6) have calculated the form of the diffraction pattern to be expected from an octahedral crystal, and on the basis of the Stachel theory Laue(7) has interpreted the spot positions in transmission patterns obtained by Brück(4) from silver, nickel and cobalt condensed on heated rocksalt cleavage faces, and also in Cochrane's(8) reflection patterns from cobalt and nickel layers of up to several thousand angstroms in thickness electrodeposited on an etched (110) face of a copper single crystal. According to this interpretation, the films condensed from the vapour, and the electrodeposited cobalt, showed development of all four types of octahedral faces, while in the electrodeposited nickel layers two of these faces preponderated, evidently yielding a sort of corrugated form of surface. A further feature of Cochrane's results was the evidence of a small pseudomorphic increase of lattice constant of thin nickel films electrodeposited on to a copper single-crystal face.

Later Kirchner and Rüdiger(9) observed that a silver film about 150 A. in thickness condensed on a heated rocksalt cleavage face yielded reflection patterns which contained abnormal spots, while a thicker film (about 900 A.) yielded patterns in which only normal spots were present, but these were elongated and displaced as a result of refraction at the film surface, the direction of displacement corresponding to a cube-face boundary parallel to the substrate surface. Refraction effects attributable to octahedral boundary planes were absent, even when the beam was parallel to such planes. This observation, together with the previous experimental data(3) on change of pattern with film-thickness, suggests that the films tend to revert to normal structure with cube-face boundary as the thickness of the deposit increases. As Laue(10) pointed out, this does not affect the fact that his theory of the dependence of the diffraction patterns on crystal form did explain many of the patterns then published, and this interpretation appeared to be more satisfactory than the postulate of a cross-grating action of the planes most densely filled with atoms, although the theoretical positions of the spot diffractions would be the same in the two cases and in agreement with the observed patterns.

It seems probable that no marked octahedral-plane refraction effect would be expected in the spot pattern, even in the most favourable case, i.e. with the electron beam nearly parallel to octahedral planes, until these are well developed both in perfection and extent as a result of growth; and before this has occurred the cube-face boundary parallel to the substrate appears to predominate, as Kirchner's photographs indicate. A possible picture of the conditions from this point of view would thus be that initial nuclei of the deposit grow in size as deposition proceeds, perhaps developing octahedral faces which soon meet those of neighbouring nuclei, whereupon the subsequent deposition will add to the film-thickness without resulting in any further increase in size of the octahedral projections; and, since in the present case the alignment is practically perfect with respect to the rocksalt substrate lattice, the separate octahedral projections might well be supposed to grow together and tend to develop eventually a single external boundary face parallel to the substrate as deposition of further atoms proceeds.

The abnormal diffraction spots and lines which appear at certain definite positions simply related to those of the normal spot pattern in some of Cochrane's patterns from electrodeposited nickel films were interpreted by him(8) as due to repeated twinning of the structures on the octahedral planes, the spots resulting from the twinned layers of considerable thickness parallel to these planes and the lines from a displacement of the junction planes between, and parallel to, the twin-layers. Menzer(11, 12) has discussed in detail the reciprocal lattice and diffraction pattern of twinned spherical-close-packing structures and of regular hybrid variations of the cubic and hexagonal densest packings of spheres built up by superposition of plane hexagonal networks of atoms in definite sequence. He concluded(13) that Brück's silver and nickel films grown on rocksalt were not uniform single crystals with axes parallel to those of the rocksalt, but consisted of (111) layers of atoms interpenetrating in the same orientation as the lattices twinned on the {111} planes of a hypothetical metal lattice parallel to that of the rocksalt, and he found that the mode of orientation of such silver and nickel lattices on rocksalt and of rocksalt deposited on the silver films was then in accordance with the well-established fact that when one lattice grows on another in a definite relative orientation the boundary planes of the two lattices at the junction contain parallel atom rows of closely similar spacing. Since the observed strong satellite diffraction spots correspond to *all* the points which divide into six equal parts the diagonals of the body-centred cubic identity cells in the reciprocal lattice of the above hypothetical silver lattice, he inferred that the interpenetration of the layers occurs to such an extent that the small disturbed regions which occur at the junctions between such non-parallel twin layers form a considerable part of the film. Thus, while the above four twins on {111} would account for only some of these points, the buffer lattice between two twin layers which are parallel to two octahedral faces touching only at an octahedral edge lying in a cube face forms a lattice identical with the initial hypothetical lattice parallel to that of the rocksalt, together with a junction region, between this and the twin lattices, in which there is an incomplete filling up of the {111} planes. The latter have therefore a true periodicity of three times the normal spacing (every third plane continues undisturbed through the twin lattice and the buffer lattice), and the reciprocal lattice of this region has an identity cell which is the same as that of the hypothetical initial lattice, with the addition of the whole set of points on the diagonals mentioned above.

It has already been mentioned that besides the strong satellite spots described above as due to geometrical twin crystals and their interpenetration regions, there occur spots which correspond to the literally irrational points of intersection of the sphere of reflection with the identity-cell diagonals of the initial lattice parallel to that of the substrate. These were interpreted by Menzer in terms of Laue's spur theory as due to the formation of external faces. He showed that this theory is in good agreement with the spot-positions in Brück's transmission photographs from silver and nickel, the [001] spurs being most strongly developed; i.e. the films are bounded principally by an almost plane surface parallel to the substrate surface—a (001) face—in agreement with Kirchner and Rüdiger's contention, and to a lesser extent by octahedral faces. Menzer cited the appearance of only two instead of four satellite spots in one of Cochrane's photographs from a nickel film as indirect confirmation of the correctness of the interpretation of the patterns in terms of the *Stachel* theory, showing marked development of only two octahedral faces; and he cited the same evidence against Kirchner and Lassen's(3) explanation in terms of a pseudo-cross-grating diffraction from the {111} planes arising through their being the planes with the densest packing of the face-centred-cubic structure.

The continuous diffraction lines which occur besides the twin spots in one of Cochrane's nickel patterns† and which correspond to the nearly tangential intersections of two reciprocal-lattice cell diagonals with the sphere of reflection, were apparently considered by Menzer as falling outside the theory of *Stacheln* in view of boundary faces. He rejected Cochrane's postulate of a sideways shift of the {111} planes of atoms between two superposed twin layers causing a cross-grating type of diffraction from these planes, and pointed out that such an effect might result from an aperiodic lattice disturbance arising from the slight difference between the distance apart of the superposed hexagonal atomic networks when they occur in the cubic and hexagonal close-packing arrangements respectively, a difference which is indicated by the known lattice constants of the cubic and hexagonal close-packing forms of nickel. This interpretation seems to be equally applicable to the other cases in which the cell diagonals intersect the sphere of reflection at a larger angle, and which were interpreted as showing development of boundary faces.

Kirchner and Cramer(14) have recently expressed agreement with the conception of a {111} twinned structure of the metal films, and have obtained a pattern similar to those from silver on rocksalt from the etched surface of a natural copper crystal which was thus presumably actually twinned submicroscopically on {111} analogous to the macroscopic twinning on {111} of copper and gold already observed by mineralogists. They obtained patterns from films of silver of various thicknesses, condensed on rocksalt, and showed that in the patterns from the films of only about 8 A. (or two unit cells) thickness the twin spots were hardly observable, while the normal spots were broadened and displaced due to refraction at the boundary face parallel to the substrate surface and were connected by broad lines similar to those obtained by Cochrane. It may be noted that there are no signs of spots due to the rocksalt substrate in the pattern reproduced, though such spots would certainly be expected for a film-thickness as small as that stated, or even considerably greater than it, at the angle of grazing incidence used. Films about 20 A. thick showed sharper diffraction spots and noticeable formation of satellite spots, and the latter were strongly developed in patterns from films about 200 A. thick, while with still thicker films the satellite spots became relatively weaker and were not obtained from films about 1000 A. thickness.

---

† Reference (8), figure 6.

It may further be mentioned that the bands and extra rings observed(15, 16) in patterns from polycrystalline films of silver, nickel, cobalt and some other face-centred cubic metals deposited from the vapour or electrolytically on inert or polycrystalline substrates were interpreted by Quarrell(16) as evidence of a transition from an initial close-packed hexagonal structure such as cobalt and nickel are known to form under certain conditions, to the normal face-centred cubic close-packing structure. Here, evidently, as Quarrell suggested, the metal atoms first deposited on a neutral plane surface such as cellulose pack themselves together in a plane hexagonal network; and successive layers may be built up either in the cubic or hexagonal close-packing arrangement, resulting in a hybrid close-packed lattice of the type discussed by Menzer. The close similarity between such abnormal diffractions and those obtained from silver condensed on rocksalt was, indeed, pointed out by Finch and Wilman(17), who showed that the bands in transmission patterns from silver electrodeposited on to a polycrystalline metallic substrate could be resolved into crossed lines which became enhanced into spots on the 111, 200, etc., rings, closely analogous to the crossed streaks and satellites of the patterns from Kirchner's and Brück's silver films.

An account will now be given of a further study of the electron-diffraction patterns yielded by silver films and of the effect of heat on these films.

§ 2. EXPERIMENTAL PROCEDURE

Silver films were prepared by condensation from the vapour in vacuo on to fresh cleavage surfaces of good rocksalt crystals maintained at a temperature of about 200° C.

Films of suitable thickness were obtained by vaporization of the silver, during about fifteen minutes, from the wire heated electrically to bright red heat at a distance of approximately 1·5 cm. from the rocksalt face. These films were still partly transparent and dark blue in transmitted light, though they had good metallic reflecting power. Electron-diffraction patterns from the films were photographed before and after removal of the rocksalt by solution. The Finch(17) type of diffraction camera was used, the camera-length and voltage being respectively about 47 cm. and 60 kv. in almost all cases.

§ 3. THE REFLECTION PATTERNS FROM THE SILVER FILMS ON ROCKSALT

Owing to the fact that the removal of the films from their substrates inevitably introduces at least slight distortion of the film, and possibly some structural change, it seemed desirable first of all to compare the reflection patterns with those to be expected from Menzer's view of the structure of the films, since Menzer has considered only one of Brück's transmission patterns. The silver films, prepared as described above, yielded electron-diffraction reflection patterns similar to those already published.† These patterns may be interpreted in the usual way by the Laue-zone method, by which the actual pattern can be mapped out as a plane index-field in which each point is associated with the three Laue indices; these, of course, are not in general integers. The indices h, k, l, referred to the axes of a normal cubic silver lattice parallel to the rocksalt lattice, can be ascribed to the diffraction spots most conveniently by drawing the nearly straight and equidistant Laue-zones which correspond to the two main lattice rows nearest normal to the beam, and the circular zones due to the row parallel to the beam or nearly so. Owing to the limited penetration of the electron beam and the relatively small size of the crystal projections in the surface layer from which the coherent diffraction arises, the latter zones are considerably broadened and each extends over many cross-grating spots which lie at the intersections of the first two sets of Laue zones, the intensity of the spots decreasing to zero on each side of the Laue-zone median as the corresponding Laue index diverges from the integral value. It is usually convenient to characterize the diffractions by the integral order or order-number of the Laue zone on which they lie, instead of by the actual Laue index (not in general an integer) with respect to the related lattice row, which is associated with any particular direction of diffraction.

In the alternative reciprocal-lattice representation of the directions of diffraction, which, owing to its three-dimensional nature, is usually better suited to numerical and geometrical considerations than to direct comparison with the pattern, the lattice points, as Laue has shown, must be supposed extended in the directions normal to the boundary faces of the crystal. The diffraction spots lying on the Laue zone of h'th order due to the atom row $[uvw]$ correspond to intersections of the sphere of reflection with the intensity regions of a coplanar set of lattice points of the reciprocal lattice. These lattice points lie on the h'th plane from the origin which can be drawn through the lattice points of the reciprocal lattice normal to $[uvw]$, the indices‡ of the plane being therefore $(uvw)^*$, since for all diffractions hkl which lie on that Laue zone we have $h' = uh + vk + wl$, and this is the condition that the point $[[hkl]]^*$ shall lie on the h'th $(uvw)^*$ plane from the origin.

In the case of a crystal oriented with a main atom row parallel to the beam, the diffraction-spot positions are thus, for the small angles of diffraction considered, very nearly the same as those obtained by imagining the reciprocal lattice set up in the corresponding orientation with its origin at the central spot and enlarged to λL times its dimensions, and then the successive planes of lattice points normal to the beam-direction projected on to the corresponding circular Laue-zone regions, drawn on the pattern in the usual way, for the atom row which is parallel to the beam.§ Figure 1, derived from Menzer's diagram,|| shows the identity units in some of the successive reciprocal lattice planes normal to the beam for the four $\{111\}$

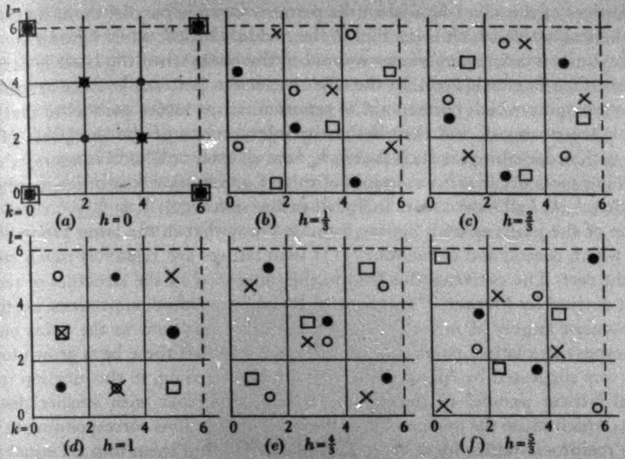
---

† Reference (17), figures 60 and 61.

‡ Asterisked symbols refer throughout to the reciprocal lattice.

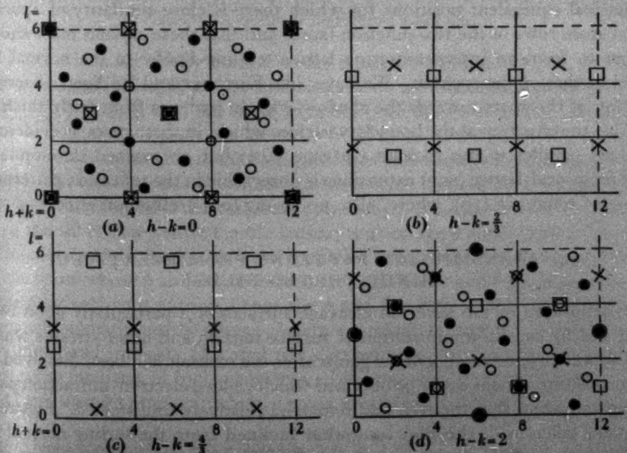
§ This is, in effect, the method which Menzer adopted(13).

|| Reference (12), p. 395.



(*a*) $h=0$ (*b*) $h=\frac{1}{3}$ (*c*) $h=\frac{2}{3}$ (*d*) $h=1$ (*e*) $h=\frac{4}{3}$ (*f*) $h=\frac{5}{3}$

Figure 1. Identity units of the combined reciprocal lattice of the four silver lattices which are in $\{111\}$ twin positions in relation to a lattice parallel to the rocksalt lattice, in successive planes ($h=0, \frac{1}{3}, \frac{2}{3}$, etc.) normal to the *a* axis of the latter. The co-ordinates of the points in the lattice reciprocal to the initial lattice are *h*, *k*, *l*. The reciprocal lattice points associated with the various twins are distinguished by the following symbols: ●, twin on $[111]$; ×, twin on $[\bar{1}11]$; ○, twin on $[\bar{1}\bar{1}1]$; □, twin on $[1\bar{1}1]$. Points of the reciprocal lattice due to the silver lattice parallel to that of the rocksalt occur at all points in the diagrams which have integral *hkl* co-ordinates.



(a) $h-k=0$ (b) $h-k=\frac{2}{3}$ (c) $h-k=\frac{4}{3}$ (d) $h-k=2$

Figure 2. Identity units of the combined reciprocal lattice of the four twins on $\{111\}$, in successive planes ($h'_{[1\bar{1}0]}=h-k=0, \frac{2}{3}, \frac{4}{3}$, etc.) normal to the $[1\bar{1}0]$ axis of the initial lattice. Twin-point symbols and *hkl* co-ordinates as in Figure 1.

twin lattices of the silver film when the primary beam is parallel to the *a* axis of the hypothetical lattice which is parallel to the rocksalt lattice; while figure 2 shows the identity units in successive planes normal to the beam when the latter is parallel to the cube-face diagonal $[\bar{1}10]$. In the silver reflection patterns at these azimuths the diffraction spots which correspond to a normal silver lattice parallel to that of the rocksalt are strongest, and they lie on the intersections of the three sets of Laue zones with order-numbers such that h, k, l are all even or all odd integers, while the remaining spots occur at intersections of three Laue zones whose order-numbers are such that h, k, l all have values included in the series $\pm\frac{1}{3}, \frac{2}{3}, \frac{4}{3}, \frac{5}{3}, \frac{7}{3}, \ldots$ etc. Comparison of the patterns with figures 1 and 2 shows that in the latter group of spots those which correspond to the four $\{111\}$ twin lattices are relatively much stronger than the rest. The patterns thus lead to the same view of the structure of the films as that derived by Menzer(13) from one of Brück's transmission patterns, i.e. that the films consist largely of silver crystals in the same positions as the twins on $\{111\}$ with respect to a lattice parallel to the rocksalt, and that these have grown together in the way suggested by Menzer, thereby actually forming, in the junction regions, normal lattices parallel to that of the rocksalt, together with smaller disturbed regions which cause the appearance of the other diffractions corresponding to points on the reciprocal-lattice identity-cell diagonals dividing them into six equal parts.

We have never observed in the patterns any signs of spots which might indicate an actual twinning of each of the above four lattices on all their octahedral planes, as Kirchner and Cramer supposed to occur. Thus the formation of the silver films seems to follow the course suggested by Menzer, namely growth of crystal nuclei orientated with the $\{221\}$ plane parallel to the rocksalt cleavage face in the four symmetrical equivalent positions for which there is close similarity of spacing in parallel atom rows of the two junction faces; and that these crystals meet and grow together to form an interpenetrating lattice tending finally to the normal lattice parallel to that of the rocksalt. We have, like Kirchner and Rüdiger, observed an extension of the spots towards the shadow-edge in patterns from fairly thick silver films, due to refraction at the boundary surface which, in these cases, is evidently the cube face parallel to the rocksalt cleavage face; but no marked development of $[001]^*$ reciprocal-lattice point extensions is observable in the reflection patterns such as those of reference (17), where, also, no strong (001) refraction effect is present.

§4. TRANSMISSION PATTERNS FROM RELATIVELY UNDISTORTED FILMS

By deposition of the silver on cleavage surfaces of a particularly good rocksalt crystal and by careful minimization of surface tension and other stresses while the thin film was supported on a metal gauze after the rocksalt had been dissolved away, transmission specimens could be obtained which yielded electron-diffraction patterns of a type showing that the film had remained relatively undistorted. Figure 3 for example†, taken with the film somewhat inclined from the setting normal to the electron beam, shows such a pattern in which the absence of strong distortion can be seen at once from the fact that those satellite spots which occur in groups near a vertical line through the central spot in the pattern show a characteristically increasing divergence from each other with increasing distance from the central spot. As with the reflection patterns, the indexing is conveniently referred to the axes of the silver lattice which is parallel to the rocksalt lattice. The Laue zones due to the lattice rows parallel to those cubic axes are readily distinguishable and the projections $\bar{a}$, $\bar{b}$, $\bar{c}$ of the corresponding axial directions are indicated in the lower right-hand corner of figure 3; thus the indices *h*, *k*, *l* of the diffractions can be inferred directly in this case. The narrow and practically straight Laue zones due to the two axes *a* and *b* which are nearly normal to the beam are marked in figure 3, together with their integral indices *h*, *k*. The twin spots lie on intermediate zones with *h* and *k* equal to $\pm\frac{1}{3}$, $\frac{2}{3}$, $\frac{4}{3}$, $\frac{5}{3}$, $\frac{7}{3}$, etc., but these have not been drawn so as not to overcrowd the diagram, since their positions are clear from those of the zones with integral *h* and *k*. The curved zones due to the *c* axis, which is only slightly inclined to the beam, are broad, and each extends over many of the *a*-zone and *b*-zone intersections, the diffractions with integral order-numbers occurring in accordance with the structure factor of the silver face-centred cubic lattice; thus on the *c* zones with even order *l* the diffractions appear for which *h* and *k* also are even, and on the *c* zones with odd order *l* lie spots whose *h* and *k* indices also are odd. The *a* and *b* zones are clearly in this case relatively very narrow, hence the *h* and *k* indices of the diffractions lying on them will actually nearly always closely approach to integral values; but the actual *l* index and intensity of a spot will depend upon the relative distance of the spot from the medians, or loci of maxima, of the two neighbouring *c* zones. In the present case, owing to the twin spots, *c* zones must also be drawn for $l = \pm 0, \frac{1}{3}, \frac{2}{3}, 1, \frac{4}{3}, \frac{5}{3}, \ldots$.

---

† This and all the other photographs have been reduced to rather less than half the linear dimensions of the originals.

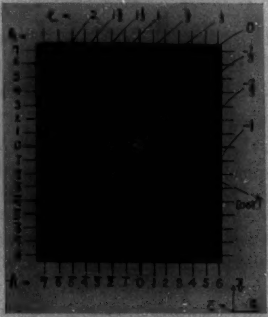


Figure 3. A relatively undistorted silver film.

The curves along which the *l* index is $\pm 0, \frac{1}{3}, \frac{2}{3}, 1$, etc. have been drawn on the pattern itself for clarity in discussion of the pattern. These curves were derived as follows. First the orientation of the lattice with respect to the primary beam was calculated from the indices of three of the most intense spots, which were assumed to be close to the maximum of the Laue zones on which they lie. It can be shown† that the condition for a diffraction with indices *hkl* to occur, i.e for the point $[[hkl]]^*$ in the reciprocal lattice to lie on the sphere of reflection, is, for the case of a cubic lattice whose reciprocal lattice constants are a^*, b^*, c^*,

$$Uh+Vk+Wl=\tfrac{1}{2}(h^2+k^2+l^2), \qquad \text{......(1)}$$

where U, V, W are the a^*, b^*, c^* coordinates of the centre of the sphere of reflection in the reciprocal lattice. Allowing for the radial decrease in intensity due to the atom-form factor, three suitable well-defined spots are those with indices $\bar{1}71$, $\bar{5}\bar{1}1$, $5\bar{3}\bar{1}$. Three equations are obtained by substitution of these indices in equation (1), yielding $[[U, V, W]]^*=[[\frac{37}{2}, -\frac{31}{4}, \frac{393}{4}]]^*$. The angles θ_a, θ_b, θ_c made by the beam with the *a*, *b*, and *c* axes respectively are given by

$$\cos\theta_a=U/\sqrt{\Sigma U^2}, \quad \cos\theta_b=V/\sqrt{\Sigma U^2}, \quad \cos\theta_c=W/\sqrt{\Sigma U^2}; \qquad \text{......(2)}$$

thus in the present case $\theta_a=79°\ 22'$, $\theta_b=180°-85°\ 34'$, $\theta_c=11°\ 36'$. The projections $\bar{a}$, $\bar{b}$, of the *a* and *b* axes on the screen, which is normal to the beam, are obtained by drawing perpendiculars from the central spot to the related Laue zones, but the $\bar{c}$ direction must be calculated from the equation

$$\cos\beta=-\cot\theta_a\cot\theta_b, \qquad \text{......(3)}$$

whence β, the angle between the $\bar{a}$ and $\bar{c}$ projections, is $180°-23°\ 50'$. Similarly the angle between $\bar{a}$ and $\bar{b}$ is $\bar{\gamma}=180°-89°\ 10'$. Since the camera length L was 47·0 cm., the *c* axis produced backwards must meet the screen at a distance R_0 equal to 47·0 tan 23° 50′ or 9·65 cm. from the central spot along the $-\bar{c}$ direction; and since θ_c is small, the *c* zones are practically circles over the range of scattering angles considered, with this point as centre and radii calculated from the Laue equation. The semi-apical angle θ_l of the *l*th-order Laue cone about the *c* axis is given by

$$\cos\theta_l=\cos\theta_c-l\lambda/a. \qquad \text{......(4)}$$

In the present case $\theta_c=11°\ 36'$; and the *a* zone spacing, which is obtained from the distance apart of two strong and well-defined spots on the same *b* zone but different *c* zones in order to avoid possible spot-displacement effects, is found to be 0·557 cm.; hence we have

$$\lambda/a=(0{\cdot}557\sin\theta_a\sin\bar{\gamma})/L=0{\cdot}01165.$$

Thus, from equation (4), $\cos\theta_l=0{\cdot}9795-0{\cdot}01165\,l$,

and the radius R_0 of the corresponding Laue zone, taken as circular with centre at the point of intersection of the *c* axis with the screen, is given by

$$R_l=R_0+L\tan(\theta_l-\theta_c), \qquad \text{......(5)}$$

where $R_0=L\tan\theta_c=9{\cdot}65$ cm. The radii of the zones with l equal to 2, $\frac{5}{3}$, $\frac{4}{3}$, 1, $\frac{2}{3}$, $\frac{1}{3}$, 0, $-\frac{1}{3}$, $-\frac{2}{3}$ and -1 are thus calculated to be 14·11, 13·47, 12·78, 12·05, 11·35, 10·51, 9·65, 8·71, 7·63 and 6·43 cm. respectively.

---

† Reference (18), equation (14).

Since the beam is inclined at a relatively small angle to the *c* axis, figure 1 can be used also in the present case to distinguish the twin spots from those ascribed to their junction regions, if we replace *k*, *l* and *h* in figure 1 by *h*, *k* and *l* respectively, and if we omit the distinction between the spots from the different twins. Thus, in considering spots lying on the *l*th-order zone in figure 3, it is only necessary to compare their positions relative to the even-order *h* and *k* zones, with the twin-spot positions in figure 1 relative to the corresponding lines along which *h* and *k* are 0, 2, 4, 6, etc., in the diagram of the reciprocal-lattice plane having the *l* value considered. It will be seen that the strongest twin spots lie very close to the circles for which *l* is exactly $\pm\frac{1}{3}, \frac{2}{3}, \frac{4}{3}, \frac{5}{3}$, etc., for example the diffractions whose Laue-zone order-numbers are $\frac{\bar{5}}{3}\,\frac{\bar{1}}{3}\,\frac{1}{3}$, $\frac{\bar{5}}{3}\,\frac{1}{3}\,\frac{1}{3}$ and $\frac{7}{3}\,\frac{\bar{7}}{3}\,\frac{1}{3}$; while twin spots are visible even if they occur at a considerable distance from the maxima, such as $\frac{\bar{8}}{3}\,\frac{2}{3}\,\frac{2}{3}$, $\frac{\overline{10}}{3}\,\frac{2}{3}\,\frac{2}{3}$, $\frac{7}{3}\,\frac{7}{3}\,\frac{\bar{1}}{3}$ and $\frac{11}{3}\,\frac{7}{3}\,\frac{\bar{1}}{3}$. The spots ascribed to the junction regions are relatively few and weak, only occurring with moderate intensity close to the *c* zone maxima, for example the spots $\frac{1}{3}\,\frac{11}{3}\,\frac{1}{3}$ and $\frac{10}{3}\,\frac{\bar{1}}{3}\,\frac{\bar{8}}{3}$ are among the strongest of these to be found in the pattern. The pattern thus shows general agreement with the reflection patterns and with the transmission pattern analysed by Menzer. A large breadth of the *c* Laue zones is compatible with the thinness of the film from which the pattern was obtained, but the fact that the overlapping of the *c* zones of spots is so considerable as to be hardly ascribable entirely to dynamic interaction seems to indicate a slight distortion of the film, as also does the appearance of the faint spots which have apparently the Laue-zone order numbers $\frac{13}{3}\,\frac{1}{3}\,\frac{\bar{7}}{3}$, $\frac{13}{3}\,\frac{\bar{1}}{3}\,\frac{\bar{7}}{3}$ and the pair diametrically opposite, which are probably actually due to the twin spots which have these *h* and *k* indices but $l = \pm\frac{1}{3}$ instead of $\frac{2}{3}$, which may occur as a result of combined film thinness along [001] and slight distortion. This is also similar to Menzer's findings in his analysis of Brück's silver transmission pattern, and indeed some distortion would always be expected owing to the inherent difficulty of mounting for transmission such thin films.

§5. TRANSMISSION PATTERNS FROM DISTORTED FILMS

Considerable distortion of the silver films sometimes occurred during the dissolution of the rocksalt substrate and the mounting of the film on metal gauze, especially with the thinner films and if the substrate rocksalt cleavage face was not very perfect. A slight sagging and curvature of the film in the meshes of the supporting gauze is normally detectable in all transmission specimens, but the patterns from some films indicated a rather high degree of distortion. Figure 4 shows such a pattern, obtained at practically normal incidence of the beam to the mean film plane. An outstanding feature is that the many diffractions which are present lie on definite loci, some of which themselves are visible as continuous lines in the pattern; hence it is at once clear that the film-distortion has not been a random one but has taken the form of either local or general curvature of the film about definite axes. The pattern is thus the same as that which would be yielded by an undistorted film which had been rotated about each of these axes in turn during exposure of the plate, and the formation of the diffraction lines which join up the spot diffractions must be attributed to the crystal-shape and twin-plane cross-grating effects to which correspond the lattice-point extensions or spurs in the reciprocal lattice. The positions of the continuous diffraction lines and of the superposed spots provides a wealth of detail from which can be determined the axes of curvature, the extent of curvature and the relative development of the reciprocal lattice-point extensions.

The *hk*o spots in figure 4 are strongest, which is consistent with the orientation having been such that the mean film plane was practically normal to the beam. The intensities of the other diffractions indicate the presence of a distortional deviation from this orientation in parts of the film, whose relative extent of occurrence in the film decreased the larger the amount of the deviation. It is most convenient to refer the positions of diffractions in the pattern to those of the *hk*o spots which would be yielded by an undistorted film in this main setting normal to the beam, and to consider the effect of rotation of an undistorted film about a given axis from this initial setting.

Figure 4. A much distorted silver film, nearly normal to the beam.

In the case of figures 4, 8, 10, 11 and 13, however, measurement of the radial distances of the spots showed that these spots were definitely in the silver 110 positions and were not due to sodium chloride.

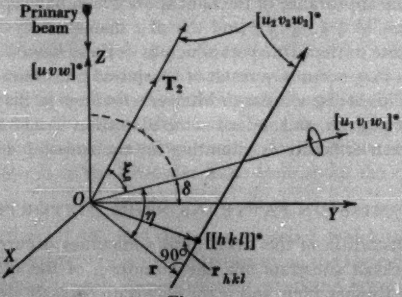


Figure 5.

General form of the rotation pattern. The form of the electron-diffraction pattern produced by rotation of a crystal about any given axis has already been discussed(19). It has been shown that if the rotation axis is inclined at $(90^\circ - \delta)$ to the beam direction, figure 5, and if three orthogonal axes *OXYZ* are chosen so that *OZ* lies along the beam direction and the plane OZY contains the rotation axis, then a reciprocal lattice row which is inclined at ξ to the rotation axis and whose normal from the origin is of length r and is inclined at η to the rotation axis, intersects the XY plane in the locus

$$x^2 \cos^2 \xi + y^2 (\sin^2 \delta - \sin^2 \xi) + 2yr \cos \eta \cos \delta - r^2 (\cos^2 \xi + \cos^2 \eta) = 0. \quad \ldots\ldots(6)$$

Since the XY plane is normal to the beam it approximates closely to the part of the sphere of reflection with which we are concerned, and this locus has the same geometrical form as the corresponding locus in the diffraction pattern itself which is λL times as large. If the reciprocal lattice contains rows which have a very marked high concentration of lattice points, or if, as in the present case, there are strongly developed lattice-point extensions in certain directions, the corresponding loci of point groups or continuous lines respectively will be prominent in the pattern. The centre of the conic represented by equation (6) always lies on OY, i.e. on the projection of the rotation axis on the XY plane which is normal to the beam, at the point O_1, i.e. (x_1, y_1), where

$$x_1 = 0, \quad y_1 = -r \cos \eta \cos \delta / (\sin^2 \delta - \sin^2 \xi) \quad \ldots\ldots(7)$$

and its equation with respect to parallel axes X_c, Y_c through its centre as origin is

$$x_c^2 + y_c^2 (\sin^2 \delta - \sin^2 \xi)/\cos^2 \xi = r^2 \{1 + \cos^2 \eta / (\sin^2 \delta - \sin^2 \xi)\}. \quad \ldots\ldots(8)$$

The axes of the conic are parallel to the X and Y axes, and if $r = 0$ the locus can only reduce to the pair of lines

$$y = \pm x \cos \xi / (\sin^2 \xi - \sin^2 \delta)^{\frac{1}{2}}, \quad \ldots\ldots(9)$$

but if $r \neq 0$ the locus is a hyperbola, a parabola or an ellipse according to whether $|\xi|$ is $>$, $=$ or $< |\delta|$. The asymptotes of the hyperbolas are the lines

$$y_c = \pm x_c \cos \xi / (\sin^2 \xi - \sin^2 \delta)^{\frac{1}{2}}, \quad \ldots\ldots(10)$$

through the centre. For the special case when $|\xi| = |\delta|$ the equation of the parabola with respect to the original axes X, Y is, from (6),

$$x^2 = -2yr \cos \eta / \cos \delta + r^2 (1 + \cos^2 \eta / \cos^2 \delta), \quad \ldots\ldots(11)$$

thus the axis of the parabola is OY and its vertex is at the point

$$(0, \{\cos \delta / \cos \eta + 1\} . r/2).$$

In the present case of a cubic lattice we may first consider, quite generally, the locus obtained when the crystal is set up initially with the lattice direction $[uvw]$ parallel to the reversed beam direction, then rotated about $[u_1 v_1 w_1]$ (the indices of these directions are the same in the reciprocal lattice for the cubic crystal) so that the locus is traced out in the XY plane normal to the beam by the reciprocal-lattice line whose direction is $[u_2 v_2 w_2]^*$ and which passes through the point $[[hkl]]^*$ in the reciprocal lattice. Since $(90° - \delta)$ is the angle between $[uvw]^*$ and $[u_1 v_1 w_1]^*$, and ξ is the angle between $[u_1 v_1 w_1]^*$ and $[u_2 v_2 w_2]^*$, we have

$$\sin \delta = (uu_1 + vv_1 + ww_1) / \sqrt{\Sigma u^2} . \sqrt{\Sigma u_1^2}, \quad \ldots\ldots(12)$$

$$\cos \xi = (u_1 u_2 + v_1 v_2 + w_1 w_2) / \sqrt{\Sigma u_1^2} . \sqrt{\Sigma u_2^2}. \quad \ldots\ldots(13)$$

If $\mathbf{r}_{hkl}$ represents the radius vector from the origin to the point $[[hkl]]^*$ and $\mathbf{r}$ is the radius vector normal to the line $[u_2v_2w_2]^*$ through $[[hkl]]^*$, then

$$\mathbf{r} = \mathbf{r}_{hkl} - \{(\mathbf{r}_{hkl} . \mathbf{T}_2)/T_2\}\, \mathbf{T}_2/T_2,$$

where $\mathbf{r}_{hkl} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$, and $\mathbf{T}_2 = u_2\mathbf{a}^* + v_2\mathbf{b}^* + w_2\mathbf{c}^*$; thus

$$\begin{aligned}\mathbf{r} &= \mathbf{r}_{hkl} - \{(hu_2 + kv_2 + lw_2)/\Sigma u_2^2\}\, \mathbf{T}_2 \\ &= (h - Mu_2)\, \mathbf{a}^* + (k - Mv_2)\, \mathbf{b}^* + (l - Mw_2)\, \mathbf{c}^*, \qquad \text{......(14)}\end{aligned}$$

where

$$M = (hu_2 + kv_2 + lw_2)/\Sigma u_2^2. \qquad \text{......(15)}$$

The angle η between $\mathbf{r}$ and $[u_1v_1w_1]^*$ is therefore given by

$$\begin{aligned}\cos\eta &= \Sigma\,\{u_1\,(h - Mu_2)\}/\sqrt{\Sigma u_1^2} . \sqrt{\{\Sigma\,(h - Mu_2)^2\}} \\ &= \{(hu_1 + kv_1 + lw_1) - M\,(u_1u_2 + v_1v_2 + w_1w_2)\}/\sqrt{\Sigma u_1^2} . \sqrt{\{\Sigma h^2 - M^2\Sigma u_2^2\}}. \qquad \text{......(16)}\end{aligned}$$

The length r can either be obtained from the above expression for $\mathbf{r}$, since $r = \sqrt{(\mathbf{r} . \mathbf{r})}$, or directly from the relation

$$r = r_{hkl} \sin \{[hkl]^*, [u_2v_2w_2]^*\},$$

thus

$$r = a^* . \sqrt{(\Sigma h^2 - M^2\Sigma u_2^2)}, \qquad \text{......(17)}$$

where a^* is the length of the reciprocal axes of the cubic lattice. In the equations relating to the positions of the locus we need r and $r\cos\eta$ but not $\cos\eta$ alone, thus it is convenient to use the expression

$$r\cos\eta = a^* . \{\Sigma hu_1 - M\Sigma u_1u_2\}/\sqrt{\Sigma u_1^2}. \qquad \text{......(18)}$$

Comparison of figure 4 with the pattern due to bending about a cube axis in the film plane. The *hk*o spots in figure 4 may be indexed with reference to the same $\bar{a}$ and $\bar{b}$ axes as were chosen in figure 3, and will then have approximately the same disposition as in the latter. These spots correspond to the reciprocal lattice points in the a^*b^* plane, which is nearly normal to the beam in the main orientation, as is represented in figure 6. Parts of hyperbolic curves, intensified where they cross normal Hull-Debye-Scherrer ring positions, can be seen in figure 4 (and more clearly in the original plate), having axes parallel to $\bar{a}$ and $\bar{b}$; for example the curve which passes through the $\bar{2}20$ spot towards the central spot but curves upward and passes through the 020 spot, and that which curves the opposite way and passes through the $\bar{2}00$ spot. The related axes of bending of the film must therefore lie in the *ac* and *bc* planes respectively, and the most probable directions will be those parallel to the film plane, i.e. the *a* and *b* axes. These hyperbolae could not correspond to loci traced out in the plane normal to the beam by $[001]^*$ lattice-point extensions during rotation of the reciprocal lattice about one of these axes, since those loci would clearly be straight lines normal to the projection of the rotation axis on this plane. No clear diffraction lines of this type are to be observed in the pattern; thus it appears that in this case, as with the reflection patterns such as those of reference (17), the $[001]^*$ lattice-point extensions were not appreciably developed. On the other hand, the above hyperbolas do agree well in position with those corresponding to the cell-diagonals in figure 6, for rotation axes along a^* or b^*. Thus, consider the loci due to the lattice lines $[11\bar{1}]^*$, the rotation axis being along a^*. In this case we have

$$[uvw]^*=[001]^*; \quad [u_1v_1w_1]^*=[100]^*; \quad [u_2v_2w_2]^*=[11\bar{1}]^*; \qquad \text{......(19)}$$

and let $[[hkl]]^*$ be still unspecified. Then from (12), (13), (17) and (18) we have

$$\left.\begin{aligned} \sin\delta=0; \quad \cos\xi=1/\sqrt{3}, \quad \sin\xi=\sqrt{\tfrac{2}{3}}; \\ r=a^*.\sqrt{\tfrac{2}{3}}.\{h^2+k^2+l^2-\tfrac{2}{3}(hk-kl-lh)\}; \\ r\cos\eta=a^*.(2h-k+l)/3. \end{aligned}\right\} \qquad \text{......(20)}$$

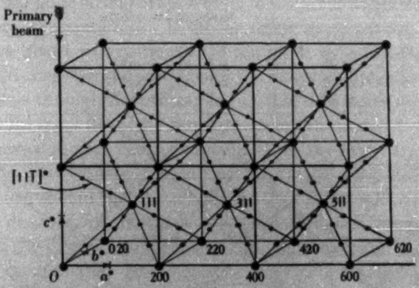


Figure 6. Perspective view of part of the reciprocal lattice of the silver films.

Since $|\xi|$ is $>|\delta|$ the locus is a hyperbola, and with the X and Y axes as defined previously the centre of the conic is, by equations (7), at

$$x_1=0; \quad y_1=a^*.(2h-k+l)/2, \qquad \text{......(21)}$$

and the equation of the hyperbola with respect to parallel axes through this point is, by equation (8),

$$x_c^2-2y_c^2=a^{*2}.(k+l)^2/2. \qquad \text{......(22)}$$

It will be seen from equation (22) that similar hyperbolas are associated with points which have the same $(k+l)$ values and any h value and that the hyperbolas all have the same equation for their asymptotes, referred to the X_cY_c axes. Since all the $[11\bar{1}]^*$ lines pass through $[[hk0]]^*$ points, it is sufficient to calculate the details of the hyperbolas for $[[hk0]]^*$ points, in order to derive those of the whole set. The data relating to typical hyperbolas are given in table 1, and figure 7 shows the positions of the hyperbolas near the origin. The perpendicular straight full lines in figure 7 represent the lines through the $[[hk0]]^*$ points parallel to the initial a^* and b^* directions before rotation, and the broken straight lines represent the asymptotes of the $[11\bar{1}]^*-[[020]]^*$ hyperbola taken as an example of those for the whole set.

In order to make still more clear the comparison of figure 7 with the pattern (figure 4) circles have been drawn in figure 7 to represent the inner of the Hull-Debye-Scherrer rings seen faintly in the pattern. The normal spots appearing as the result of crystal rotation about the *a* axis must, of course, lie on these; and all spots having the same *h* index must lie on the same layer line. The layer lines are practically straight and normal to the axis of rotation a^*, along which in figure 7, accordingly, has been placed the scale of *h* indices of the spots.

Table 1. Details of the positions of the hyperbolas for initial setting $[uvw]^* = [001]^*$, rotation axis $[u_1v_1w_1]^* = [110]^*$, and lattice lines $[u_2v_2w_2]^* = [11\bar{1}]^*$ passing through the points $[[hk0]]^*$

*hk*o	Coordinates (x_1y_1) of the centre	Equation of the hyperbola referred to the axes X_cY_c through its centre	X_cY_c coordinates of the vertex
*h*00	$0, ha^*$	$x_c^2 - 2y_c^2 = 0$	$0, 0$
*h*20	$0, (h-1)\,a^*$	$x_c^2 - 2y_c^2 = 2a^{*2}$	$\pm a^*.\sqrt{2}, 0$
*h*40	$0, (h-2)\,a^*$	$x_c^2 - 2y_c^2 = 8a^{*2}$	$\pm a^*.2\sqrt{2}, 0$
*h*60	$0, (h-3)\,a^*$	$x_c^2 - 2y_c^2 = 18a^{*2}$	$\pm a^*.3\sqrt{2}, 0$

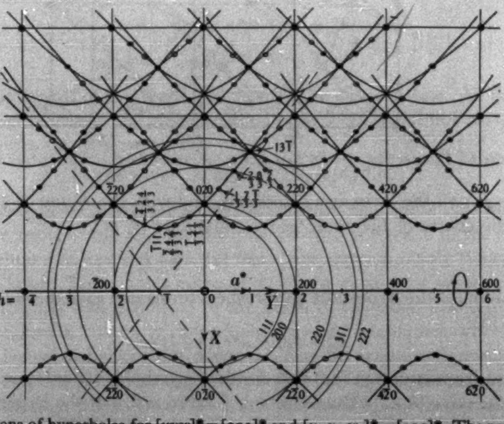


Figure 7. Positions of hyperbolas for $[uvw]^* = [001]^*$ and $[u_1\,v_1\,w_1]^* = [100]^*$. The normal *hk*o spots are represented by large black circles, other normal *hkl* spots and twin spots by smaller black circles, and junction spots by the small rings.

If the rotation were as much as 45° from the main crystal setting, lines parallel to the asymptotes of the hyperbolas would appear, passing through the *h*00 spots (*h* even) along the *Y* axis, but these are not observed in the pattern. From the extent of the innermost of those hyperbolas actually occurring noticeably and clearly visible in the pattern the amount of bending of the film about the *a* axis from the main orientation was estimated to extend to about 25°. Thus the $[11\bar{1}]^*$ hyperbola passing through the 020 point (cf. figure 7) is represented in the pattern (figure 4) only by that part of it which extends from the 020 point to near the point $\frac{\bar{1}}{3}\,\frac{4}{3}\,\frac{2}{3}$ whose

y coordinate is $-\frac{2}{3}a^*$, together with a section on the other side of the 020 spot of rather indefinite length. The $\bar{1}11$ spot, which would appear at the vertex of the hyperbola on the 111 ring, is just detectable; while the diffraction $\bar{\frac{1}{3}}\,\frac{5}{3}\,\frac{1}{3}$, which is a twin spot, is strong and occurs where the hyperbola crosses the 111 ring, in accordance with the fact that the distance of $[[\bar{\frac{1}{3}}\,\frac{5}{3}\,\frac{1}{3}]]^*$ from the origin is

$$a^*\sqrt{(\tfrac{1}{9}+\tfrac{25}{9}+\tfrac{1}{9})}=a^*\sqrt{3};$$

and the diffraction $\bar{\frac{2}{3}}\,\frac{4}{3}\,\frac{2}{3}$, which is not a twin spot, is not represented by a marked spot in the pattern. On the other side of the 020 spot the diffraction $\frac{1}{3}\,\frac{7}{3}\,\bar{\frac{1}{3}}$ (a junction spot, not a twin spot) occurs with moderate intensity and then $\frac{2}{3}\,\frac{8}{3}\,\bar{\frac{2}{3}}$ which, since it is a twin spot, is also quite strong in spite of its much lower intensity factor resulting from the following considerations. If all the orientations assumed by the crystal during its rotation about an axis normal to the beam were equally represented in the film, this would correspond to rotation of the reciprocal lattice at uniform speed about a parallel axis through the origin; and in this case a point on a given lattice line $[u_2v_2w_2]^*$ would cross the plane through the origin normal to the beam with a linear velocity normal to that plane, proportional to the perpendicular distance of the point from the rotation axis. The intensity normally associated with a point in the reciprocal lattice should then be multiplied by a factor which is inversely proportional to this distance, i.e. inversely proportional to the distance of the corresponding spots in the pattern, from the line through the central spot parallel to the projection of the rotation axis. In the present case the orientation is expected to be mainly close to that shown in figure 6, with a certain range of orientations represented in the film to a less and less extent the more the deviation from the main orientation. This would correspond to rotating the reciprocal lattice with a velocity varying with its orientation relative to the beam direction, i.e. inversely proportional to the relative extent of occurrence of that orientation in the film; and the intensity associated with points in the reciprocal lattice would then be further modified by this factor also. From the above considerations, and taking into account the radially decreasing atom-form factor, we should expect in the present case to observe only the parts of the hyperbolas (and spots lying on them) comparatively near the *hk*o spots through which they pass, the intensity of the part extending from each *hk*o point towards the vertex of the corresponding hyperbola being higher and less rapidly decreasing along the curve than that of the part on the opposite side of the *hk*o spot; and also the diffraction lines and spots forming part of the hyperbolas more distant from the central spot should be proportionally weaker in the pattern. In effect, under these conditions, the loci would be limited to slightly curved lines through the *hk*o spots, approximating closely to the diagonals of the squares in figure 7, except for the innermost more curved and stronger hyperbolic segments. The observed pattern is, in fact, of this type, as may be seen by comparison of figures 4 and 7.

Similar considerations hold for the hyperbolas traced out by the other cell-diagonal lines. The $[1\bar{1}1]^*$ lines, by symmetry, will yield the same hyperbolic curves as those of the $[11\bar{1}]^*$ lines, but the parts which occur with appreciable strength will be mirror-symmetrical to those due to $[11\bar{1}]^*$ with respect to the X axis in figure 7. The line $[111]^*$ through $[[hk0]]^*$ and the $[11\bar{1}]^*$ line through the same point are symmetrical with respect to the a^*b^* plane and therefore give rise to the same hyperbolas (and superposed spot distribution) traversed in opposite directions for a given sense of rotation of the crystal; and similarly the $[\bar{1}11]^*$ and $[1\bar{1}1]^*$ lines through a given $[[hk0]]^*$ point lead to the same hyperbolic and superposed spot positions. The case of rotation about the b axis leads, of course, to exactly similar hyperbolas in the positions derived from those described above by a 90° rotation of figure 7 about the normal to its plane through the origin.

Many specimens were examined, and yielded patterns intermediate in type between figures 3 and 4. The extent of the distortion appreciably represented in the pattern (i.e. angular variation of orientation combined with relative degree of occurrence) appeared to vary greatly, but effectively never reached the equivalent of a 45° rotation from the main orientation about a cube axis in the film plane. In figure 8 for example, the $hk0$ spots and nearest twin spots are strong while the hyperbolic diffraction lines are relatively weaker than in figure 4. The innermost hyperbolas passing through the ±200 and ±020 diffractions are sharp and clear in the negative, however, and though the curves are difficult to distinguish in the reproduction, the spots lying on them can be seen clearly, and afford some additional confirmation of the above interpretation of the main features of the pattern. Thus the intensity distribution among the normal (i.e. integral order-number) spots shows that the mean film plane must have been somewhat inclined from the setting normal to the beam (roughly about an axis near a cube-face diagonal in the film plane) and this has resulted, as expected from the above interpretation of the patterns, in the formation of hyperbolas (due to the reciprocal-lattice identity-cell diagonals as before) completed right up to their vertices, near which a 111 type of diffraction spot appears, lying on the 111 ring.

Figure 8. Slightly distorted silver film, not quite normal to the beam. Note faint but continuous hyperbolas with spots at their apices on the 111 ring position.

Comparison of figure 4 with the pattern due to bending about face-diagonals in the film plane. There are several features of figure 4 which are not explained by the above type of bending of the film. One of the most prominent is the set of straight streaks, which are strengthened slightly where they cross the 111 and 200 rings, in the positions such as are shown by the broken thicker lines in figure 9 relative to those of the *hk*0 spots. Diffraction lines such as these would appear if some parts of the film had undergone bending through 90° from the main setting, about axes parallel to the cube-face diagonals [110] and [1$\bar{1}$0] in the film plane; and such a bending, though to a much smaller extent (about 35°), would suffice to explain the appearance of strong spots on the 111 and 200 rings where these cross the diameters at 45° to the horizontal direction in figure 9.

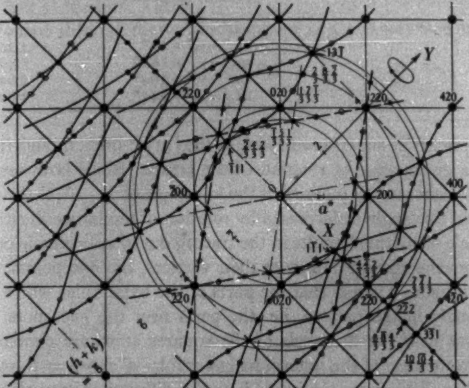


Figure 9. Positions of the hyperbolas and lines for $[uvw]^* = [001]^*$ and $[u_1 v_1 w_1]^* = [110]^*$. Spot symbols as in Figure 7.

The effect on the pattern caused by a bending of the silver films about the above axes may be illustrated by considering the case of the [110] axis of curvature. The loci traced out on the plane normal to the beam by the $[11\bar{1}]^*$ lattice line are obtained by putting in equations (12), (13), (17) and (18) the values

$$[uvw]^* = [001]^*, \quad [u_1 v_1 w_1]^* = [110]^*, \quad [u_2 v_2 w_2]^* = [11\bar{1}]^*, \qquad \text{......(23)}$$

whence

$$\left.\begin{aligned} \sin\delta = 0; \quad \cos\xi = \sqrt{\tfrac{2}{3}}; \quad \sin\xi = \sqrt{\tfrac{1}{3}}; \\ r = a^* . \sqrt{\tfrac{2}{3}} . \{(h-k)^2 + l^2 + hk + kl + lh\}^{\frac{1}{2}}; \\ r\cos\eta = a^* . (h+k+2l)/3\sqrt{2}. \end{aligned}\right\} \qquad \text{......(24)}$$

The centre of a hyperbola of this set is thus at

$$x_1 = 0, \quad y_1 = a^* . (h+k+2l)/\sqrt{2}, \qquad \text{......(25)}$$

and its equation with respect to the axes X_c, Y_c through the centre is

$$2x_c^2 - y_c^2 = a^{*2} . (h-k)^2. \qquad \text{......(26)}$$

Since all the $[11\bar{1}]^*$ lines in the lattice pass through $[[hk0]]^*$ points, all the hyperbolas are obtained by putting $l = 0$ in equations (25) and (26). Table 2 gives the data for the $[11\bar{1}]^* - [[hk0]]^*$ hyperbolas nearest to the centre, for the main $(h-k)$ values concerned, and figure 9 shows their positions relative to those of the *hk*0 points. The $[111]^*$ lines will, through symmetry, trace out the same hyperbolas (with the same

Table 2. Details of the positions of the hyperbolas for initial settings $[uvw]^* = [001]^*$, rotation axis $[u_1v_1w_1]^* = [110]^*$, and lattice lines $[u_2v_2w_2]^* = [11\bar{1}]^*$ passing through the points $[[hk0]]^*$

$(h-k)$	Coordinates (x_1y_1) of the centre	Equation of the hyperbola referred to the axes X_e, Y_e through its centre	X_e, Y_e coordinates of the vertex
0	$0,\ a^*(h+k)/\sqrt{2}$	$2x_e^2 - y_e^2 = 0$	0, 0
±2	,,	$2x_e^2 - y_e^2 = 4a^*$	$\pm 2a^*/\sqrt{2},\ 0$
±4	,,	$2x_e^2 - y_e^2 = 16a^*$	$\pm 4a^*/\sqrt{2},\ 0$
±6	,,	$2x_e^2 - y_e^2 = 36a^*$	$\pm 6a^*/\sqrt{2},\ 0$
±8	,,	$2x_e^2 - y_e^2 = 64a^*$	$\pm 8a^*/\sqrt{2},\ 0$

possible positions oft he superposed spots) as the $[11\bar{1}]^*$ lines, though in the opposite direction for a given sense of rotation of the lattice. Since both sets of lines are initially, in the main setting, inclined at the low angle of 35° 16′ to the plane normal to the beam and this angle is decreased by rotation of the lattice in either direction, a moderate angle of rotation results in the tracing out of appreciable portions of the hyperbolas from the starting points at their vertices, the *hk*0 points. Comparison of figure 9 with figure 7 shows that the parts of the hyperbolas which are expected to appear strongly in the pattern, i.e. the parts near the *hk*0 spots, have closely similar positions near the diagonals of the squares at whose corners lie the *hk*0 spots, except in the case of the innermost hyperbolas which pass through the spots ±200 and ±020. The spots superposed on the diffraction lines also have closely similar positions in the two cases, as is illustrated in figure 9, where some of those due to the $[11\bar{1}]^*$ line through $[[020]]^*$ have been indexed for comparison with the same diffractions in figure 7. The spots lie on layer lines normal to the rotation axis $[110]^*$, along which, in figure 9, has therefore been inserted a scale of $(h+k)$ for use in consideration of the spot indices. Owing to the twin lattices spots can appear on layer lines for which $(h+k) = \pm 0, \frac{2}{3}, \frac{4}{3}, 2, \frac{8}{3}, \ldots$ etc. It may be noted that the occurrence of the cube-axis bending as well as that about the [110] and $[1\bar{1}0]$ axes, is shown most clearly by the appearance of the hyperbolic curves such as those already described which pass through the $\bar{2}20$ spot towards the central spot and curve away from this direction to pass through the $\bar{2}00$ and 020 spots respectively.

For a 90° rotation of the lattice about the axis $[110]^*$ the $[11\bar{1}]^*$ and $[111]^*$ lines intersect the plane normal to the beam in the positions shown by the broken lines in figure 9, of which those passing through the (±2, ±2, 0) spots appear to be represented in the pattern (figure 4), as has already been pointed out. The enhanced regions which occur on these lines near their intersections with the 200 and 111 rings are in the positions corresponding to twin spots, some of which on one of the lines through the 220 spot are indexed in figure 9 as an example. No marked lines occur in figure 4 corresponding to those of the above set which pass through the origin in figure 9, and this is in agreement with the fact that no twin spots occur on the cubic identity-cell diagonals of their collective reciprocal lattices, whose sides are of length $6a^*$ and are parallel to those of the normal lattice.

The lattice lines $[1\bar{1}1]^*$ and $[\bar{1}11]^*$ are normal to the axis of rotation, $[110]^*$, and therefore trace out straight lines perpendicular to this axis in the plane normal to the beam. These are also shown in figure 9, but the positions of the spots lying on them have been omitted except for the twin-spot positions on the line through the origin, some of which have been marked as an illustration and for comparison with their relative prominence in the pattern, together with their indices in the case of those on the lower half of the line. Table 3 shows the indices of the twin points which lie in the $(1\bar{1}0)^*$ plane through the origin (cf. figure 2*a*), and are relatively near to both the origin and the a^*b^* plane; together with their distances (r) from the origin, which are the same as those of the corresponding intersections shown in figure 9. As would be expected, the corresponding spots appear prominently in figure 4. The many junction spots which also lie on these lines will contribute to the considerable intensity and continuity of the latter in the pattern.

Table 3

Indices *hkl*	r/a^*
$1\bar{1}1$	$\sqrt{3}=1\cdot732$
$\frac{4}{3}\frac{\bar{4}}{3}\frac{\bar{2}}{3}$	$\sqrt{4}=2\cdot000$
$2\bar{2}0$	$\sqrt{8}=2\cdot828$
$\frac{7}{3}\frac{\bar{7}}{3}\frac{1}{3}$	$\sqrt{11}=3\cdot317$
$2\bar{2}2$	$\sqrt{12}=3\cdot464$
$\frac{8}{3}\frac{\bar{8}}{3}\frac{4}{3}$	$\sqrt{16}=4\cdot000$
$3\bar{3}1$	$\sqrt{19}=4\cdot359$
$\frac{10}{3}\frac{\overline{10}}{3}\frac{4}{3}$	$\sqrt{24}=4\cdot899$
$\frac{11}{3}\frac{\overline{11}}{3}\frac{1}{3}$	$\sqrt{27}=5\cdot196$
$4\bar{4}0$	$\sqrt{32}=5\cdot657$

Unidentified diffractions. Practically all the details of the pattern, figure 4, have now been interpreted, but no satisfactory explanation has yet been found for several faint diffraction lines and spots. Thus, faint lines can be seen passing through the 200 and 020 spots along the lines joining them, on which, besides the 110 spot, lie two faint spots symmetrically disposed with respect to the 110 spot at a distance of 0·114 times the 200–020 separation on each side of it. A faint spot also occurs at a corner of the square whose other corners are the above two spots and that on the 111 ring where it crosses the radius through the 220 spot. Similar groups of spots and lines occur in the other three quadrants of the pattern. These spots and lines are not visible in all the transmission patterns recorded, and may perhaps be associated with particularly large curvature of parts of the film or, less probably, with slight tarnishing of the film during its mounting for transmission purposes. Figure 10, obtained at practically normal transmission, shows more clearly a somewhat similar but differently oriented group of faint spots and streaks associated with the 110 type of diffraction spot, although the rest of the pattern indicates a smaller degree of curvature of most of the film. Forbidden diffractions in the 110 positions have already been met with by Brück in patterns from thin silver films†, and also can be seen in figure 11 where there are not the above square groups of spots surrounding them. Brück considered these to be due to sodium chloride from the solution remaining in contact with the film after incomplete washing of the latter, crystallized on the silver with a cube-face parallel to that of the silver film surface and a cube-edge parallel to the silver cube-face diagonal in the film plane. The intensity distribution among the normal spots in figure 11 indicates that the film must have been somewhat inclined from the setting normal to the beam, and the appearance of strong twin spots and lines shows the presence of considerable curvature of the types already discussed. Here, more clearly than in figure 4, can be seen faint diffraction spots at a third of the distance from the central spot towards the ±020 spots, whose origin has also not yet been elucidated.

---

† Reference (4), figure 10.

Figure 10.

Figure 11.

Figure 10. Slightly distorted silver film normal to beam; note 110 type of diffraction and related group.

Figure 11. Slightly distorted silver film not quite normal to beam; note spots of 110 type and faint abnormal spots on the radii to the ±020 spots.

§6. THE EFFECT OF HEATING THE SILVER FILMS IN VACUO

Silver films which yielded patterns such as figures 3, 8, 10 or 11 were placed, together with their nickel gauze supports, in a glass tube, which was then evacuated by a rotary oil pump (Hyvac) and heated externally to between 500 and 600° c. for a period of about 2 min. and then allowed to cool in vacuo. The films then yielded transmission patterns of a type more closely similar to that of the reflection patterns yielded by the films before their removal from the rocksalt, indicating that the crystal structure of the films had become much more perfectly oriented. Thus the Laue zones of strong spots became well-defined and narrower, and diffuse Kikuchi lines and bands were visible in some cases such as that shown in figure 12, in which the innermost circular zone of strong spots is the zero-order zone due to the cube axis normal to the film plane, this axis being inclined at about 2° to the beam. The median regions of the Kikuchi bands visible in figure 12, due to planes of type (100) and (110), are strong absorption bands (in the negative) instead of the usual bands of enhanced intensity, somewhat analogous to the reversed bands observed by Boersch(20) from heated mica films. Such bands may perhaps be characteristic of a definite range of crystal thickness near that where the Kikuchi bands first become noticeable. The twin spots and junction spots in figure 12, lying on the zones whose order numbers are $\pm\frac{1}{3}, \frac{2}{3}, \frac{4}{3}, \frac{5}{3}$, etc., are very sharp, and both types appear to be about equally intense.

With slightly longer periods of heating of silver films the twin spots and junction spots become very weak relatively to the normal spot pattern as shown in figure 13, which was obtained from the film initially yielding figure 10, after 3 min. heating in vacuo. The faint 110 spots and a related neighbouring group can also still be seen faintly in the pattern, and faint diffuse bands can be seen passing through the normal spots parallel to the cube face diagonals of the film plane. About 4 min. heating of fresh silver films sufficed to cause complete disappearance of the satellite spots, leaving a pattern of normal type such as figure 14, consisting only of those diffractions characteristic of a normal silver lattice having the same orientation as that part of the film which was originally parallel to the rocksalt substrate lattice.

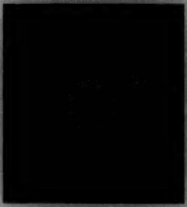


Figure 12.

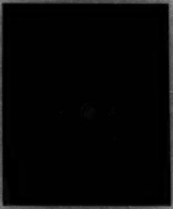


Figure 13.

Figure 12. Silver film similar to that shown in figure 3 after 2 minutes' heating in vacuo. (Camera length about half that for the other photographs.)

Figure 13. Silver film of figure 10 after 3 minutes' heating in vacuo.

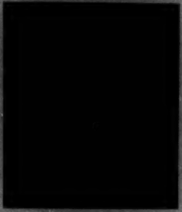


Figure 14. Silver film similar to that shown in figure 3 after 4 minutes' heating in vacuo.

The effect of heating the silver films in vacuo is thus to eliminate the discontinuities at the twin-layer boundaries and form a homogeneous single-crystal lattice parallel to that which was initially parallel to the rocksalt substrate lattice. No irrational diffractions then occur; hence even if the films were initially bounded by octahedral planes, such projections are rounded off as a result of heating, as might be expected, since a flat boundary face would provide the lowest potential energy of the crystal in view of the surface contraction (analogous to the surface tension of liquids) which Lennard-Jones$^{(21)}$ has shown should occur at the surface of ionic crystals. The final boundary faces of the silver lattice are thus evidently (001) planes, which are parallel to the film plane.

§ 7. OTHER SILVER SPECIMENS

In the course of some rough preliminary experiments, during which the temperature of the rocksalt was not regulated to the standard value used for the other silver films, a deposit was obtained which yielded the pattern, figure 15, when the beam was parallel to the cube-edge of the rocksalt substrate. In addition to the spots arranged at the corners of squares, which correspond to the usually obtained silver lattice parallel to that of the rocksalt, spots occurred in other positions. Thus, besides the 002, 004, etc. spots, the orders of the 111 diffraction of the silver occurred in the plane of incidence, hence the additional spots were due to silver crystals oriented with a {111} face parallel to the substrate rocksalt cleavage face; and since all the remaining normal diffraction spots of this group lay on faint continuous hyperbolic diffraction lines on which also lay intermediate spots corresponding to twin points on the reciprocal lattice cell diagonals, this part of the pattern was evidently due to silver crystals having practically no other restriction on their mutual orientation. If the reciprocal lattice is rotated about the $[111]^*$ axis, the asymptotes of all the hyperbolas which are traced out in a plane normal to the beam by the $[11\bar{1}]^*$, $[1\bar{1}1]^*$ and $[\bar{1}11]^*$ lattice lines enclose an angle of 38° 56′, since for these lines, by equations (12) and (13), $\sin\delta = 0$ and $\cos\xi = \pm\frac{1}{3}$, and the asymptote equations are, by (10), $y_e = \pm x_e \cot\xi = \pm x_e . \sqrt{2}/4$. The measured angles enclosed by the asymptotes of the observed hyperbolas in figure 15 are in close agreement with this value. The atoms in a (111) plane of the face-centred cubic silver lattice are arranged in a two-dimensional hexagonal network, their spacing being 2·883 A. Since this spacing is very close to half the length of the side of the rocksalt cubic cell, 5·628 A., it might have been expected that the {111} planes of the silver would have been formed in the four possible orientations such that one of their hexagon sides was parallel to one of the cube edges in the substrate rocksalt cleavage plane. The silver atoms in rows normal to these hexagon sides have a spacing of $\sqrt{3} \times 2{\cdot}883 = 4{\cdot}994$ A., i.e. the same order of magnitude as, though not very close to,

Figure 15. Abnormal pattern from silver film condensed on a heated rocksalt cleavage-face in vacuo; normal (001) orientation and also (111) orientation.

the rocksalt lattice-point spacing of 5·628 A. in this direction, which is the other cube axis direction in the rocksalt cleavage plane. No preponderance of these orientations is indicated by the diffraction pattern, however, and it appears that the above {111} orientation of the silver crystals is of the type often observed in face-centred cubic crystals deposited on flat inert substances, although the satellite spots in the pattern show that the crystals have developed strongly {111}-twinned structures of the type discussed above. We have not yet studied more closely the conditions under which such films are formed.

§ 8. CONCLUSION

This investigation has shown that the electron-diffraction patterns yielded by the silver films are in good agreement with Menzer's view of their twinned structure. The characteristic mode of distortion of the films, i.e. curvature about axes parallel to the cube axes and cube-face diagonal directions in the film surface, appears also to be compatible with the nature of the film structure and the form of the substrate surface. Thus, in common with most cleavage surfaces, that of rocksalt can often be seen to be stepped, instead of forming a perfectly continuous plane; and the electron-diffraction patterns yielded by these surfaces generally indicate a certain amount of submicroscopic roughness of the surface, presumably due to such cleavage steps. The silver deposit is therefore not to be considered as a perfectly flat film of uniform thickness, but rather of the form indicated in section by figure 16*a*. If the substrate is dissolved away and the film supported across the meshes of a metal gauze, one might expect that the resulting tension in the film would cause the angles to become more obtuse and that the parts of the film enclosing them would become curved about axes parallel to the cleavage edges, as in figure 16*b*. On this basis the curvature about axes parallel to the cube axes in the film plane, inferred from the diffraction patterns, becomes readily understandable. Whether or not such a film shape could also cause some curvature about the cube-face diagonals in the film plane in the regions near the corners of the cleavage steps is less clear. The definiteness of orientation of the axes of curvature parallel to the cube-face diagonals shown by the patterns points, rather, to their being associated with the lines of weakness along the junction regions of the {111} twin layers in the film, which also lie along octahedral edges. The question of whether the film surface is bounded by

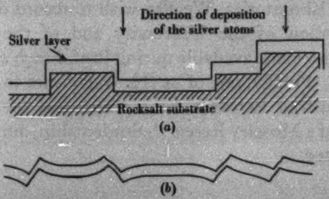


Figure 16. Diagrammatic sections of a silver film to show the probable shape of the film (*a*) as deposited on the stepped rocksalt cleavage-face, (*b*) after removal of the rocksalt.

octahedral faces, or whether the irrational diffractions are entirely due to the twinning within the octahedral-plane layers on planes parallel to the layers, seems still to be undecided; but if projections bounded by octahedral planes actually occur at the film surface, there would be lines of mechanical weakness along the direction of the troughs between them, i.e. also parallel to cube-face-diagonal directions in the film plane.

The change in crystalline structure of these silver films, observed after heating the latter in vacuo to temperatures far below the melting point, is closely connected with the experimental demonstrations of crystal growth and aggregation in metals at such temperatures, which have established the fact that the atoms at the surface of the crystals have a considerable freedom of movement over the surface of the lattice. Crystal growth in metals has been much studied by means of x-ray diffraction, and the phenomenon has recently been studied in thin metal films by the electron diffraction method(15, 22, 23, 24, 25, 26, 27, 28, 33), by microscopic examination(29, 30) and observation of optical properties(31), and also, less directly, by experiments on the temperature variation of electrical resistance of thin metal films(32, 33). A rearrangement of the silver-film structure to the extent observed in the present case at about 500° C. provides striking evidence of the high mobility of the silver atoms within the film and not merely on its surface.

§ 9. ACKNOWLEDGEMENTS

We wish to thank Prof. G. I. Finch, F.R.S., for his interest in the work, which was carried out in his laboratories. We also wish to record our indebtedness to the Conseils d'Administration of the University and of the Fondation Francqui, Brussels, for the facilities made available during Prof. Finch's tenure of the Francqui Chair during 1937–38. Finally, one of us (H. W.) wishes to thank the Department of Scientific and Industrial Research for a grant, and the Council of the Royal Society for the award of a Moseley Research Studentship, during the tenure of which this work was completed.

REFERENCES

(1) Lassen, H. *Phys. Z.* **35**, 172 (1934).
(2) Lassen, H. and Brück, L. *Ann. Phys., Lpz.*, **22**, 65 (1935).
(3) Kirchner, F. and Lassen, H. *Ann. Phys., Lpz.*, **24**, 173 (1935).
(4) Brück, L. *Ann. Phys., Lpz.*, **26**, 233 (1936).
(5) v. Laue, M. *Ann. Phys., Lpz.*, **26**, 55 (1936).
(6) v. Laue, M. and Riewe, K. H. *Z. Kristallogr*, **95**, 408 (1936).
Riewe, K. H. *Z. Kristallogr*, **96**, 85 (1937).
(7) v. Laue, M. *Ann. Phys., Lpz.*, **29**, 211 (1937).
(8) Cochrane, W. *Proc. Phys. Soc.* **48**, 723 (1936).
(9) Kirchner, F. and Rüdiger, O. *Ann. Phys., Lpz.*, **30**, 609 (1937).
Kirchner, F. *Ann. Phys., Lpz.*, **30**, 619 (1937).
(10) v. Laue, M. *Ann. Phys., Lpz.*, **30**, 616 (1937).
(11) Menzer, G. *Naturwissenschaften*, **26**, 385 (1938).
(12) Menzer, G. *Z. Kristallogr*, **99**, 378 (1938).
(13) Menzer, G. *Z. Kristallogr*, **99**, 410 (1938).

(14) Kirchner, F. and Cramer, H. *Ann. Phys., Lpz.*, **33**, 138 (1938).

(15) Finch, G. I., Quarrell, A. G. and Wilman, H. *Trans. Faraday Soc.* **31**, 1050 (1935).

(16) Quarrell, A. G. *Proc. Phys. Soc.* **49**, 279 (1937).

(17) Finch, G. I. and Wilman, H. *Ergebn. exakten Naturw.* **16**, 353 (1937).

(18) Charlesby, A., Finch, G. I. and Wilman, H. *Proc. Phys. Soc.* **51**, 479 (1939).

(19) Finch, G. I. and Wilman, H. *Trans. Faraday Soc.* **32**, 1539 (1936).

(20) Boersch, H. *Phys. Z.* **38**, 1000 (1937).

(21) Lennard-Jones, J. E. *Z. Kristallogr*, **75**, 215 (1930).

(22) Trillat, J. J. and v. Hirsch, T. *Z. Phys.* **75**, 784 (1932).

(23) Trillat, J. J. and Oketani, S. *C.R. Acad. Sci., Paris*, **203**, 1064 (1936); *J. Phys. Radium*, **8**, 59, 93 (1937).

(24) Preston, G. D. and Bircumshaw, L. L. *Phil. Mag.* **21**, 713 (1936).

(25) Finch, G. I. and Fordham, S. *J. Soc. Chem. Ind., Lond.*, **56**, 632 (1937).

(26) Clark, G. L. and Wolthuis, E. *J. Appl. Phys.* **8**, 630 (1937).

(27) Dixit, K. R. *Phil. Mag.* **16**, 1049 (1933).

(28) Rama Swamy, S. *Proc. Phys. Soc.* **46**, 739 (1934).

(29) Andrade, E. N. da C. and Martindale, J. G. *Philos. Trans.* A, **235**, 69 (1935).

(30) Essig, S. F. *J. Appl. Phys.* **10**, 61 (1939).

(31) Hass, G. *Naturwissenschaften*, **25**, 232 (1937); *Ann. Phys., Lpz.*, **31**, 245 (1938).
Rama Swamy, S. *Proc. Roy. Soc.*, A, **131**, 307 (1931); *Proc. Phys. Soc.* **46**, 739 (1934); *Proc. Ind. Acad. Sci.*, A, **1**, 347 (1934); **4**, 656 (1936).
Krautkrämer, J. *Ann. Phys., Lpz.*, **32**, 537 (1938).

(32) Lovell, A. C. B. *Proc. Roy. Soc.*, A, **157**, 311 (1936); A, **158**, 718 (1937); *Proc. Phys. Soc.* **49**, 89 (1937).

(33) Cosslett, V. E. *Proc. Phys. Soc.* **49**, 121 (1937).

SHEAR-STRESS SYSTEMS AND COMBINED STRESS

BY J. J. GUEST

Communicated by Dr W. Jevons 19 *November* 1938. *Read in title* 10 *March* 1939

ABSTRACT. This paper discusses an objection which has been raised to the law of elastic failure under combined stress which was given in a paper read before the Society in 1900—namely that, instead of the volumetric stress then used, the average volumetric stress should be employed and that only shear stresses should be associated with it. A resolution of a stress system into uniform tension and shearing stress, of a type which is advocated to-day and appears to offer advantages, is first discussed and is shown to be inapplicable as not being a complete representation. It is then pointed out that there are usually an indefinite number of shear-stress systems and only in certain cases are the shears unique. For the precise expression of a law of failure the indefiniteness must be removed, and the simple limitations available for the purpose are discussed and shown to be inapplicable.

Finally it is shown that the adoption of the suggested scheme could only be applicable to such materials as have a failure limit in simple compression less than three times that in simple tension. Thus it is incompatible with a law of failure which is to be of general application.

The results of a research described in a paper on "The strength of ductile materials under combined stress"⁽¹⁾ were there expressed as a criterion of failure as follows: "§ 56. Any state of stress can be represented by the coexistence of a shearing stress, a volumetric uniform stress, and a stress perpendicular to the plane of the shear; and the yield point occurs when the shearing stress, together with a small fraction of the volumetric stress reaches a certain amount, the volumetric stress being reckoned positive when tensile." The volumetric stress employed is $\frac{1}{2}(p_1+p_2)$, where p_1, p_3, and p_2 are the principal stresses in descending order of magnitude, and the equation of the criterion is given as

$$p_1-p_2+\lambda(p_1+p_2)=C, \qquad \ldots\ldots(1)$$

where λ varies from zero to unity with the ductility of the material.

Various criticisms have been made concerning the form of the above statement, a frequent one being that the volumetric stress to be used should be the average stress $\frac{1}{3}(p_1+p_2+p_3)$, and that shear stresses alone should be associated with it.

As the fundamental strains are those of volume and shape this criticism is logical, but the expression of the law experimentally reached in such terms and in terms of the corresponding stresses was considered and found to present difficulties. Those relating to the shear stresses are considered initially.

It will be convenient firstly to consider recent suggestions that a certain modern view of stress-resolution should be adopted. This view will be given adequately by a citation from the latest technical treatise, *Elasticity*, by R. V. Southwell, F.R.S., Professor of Engineering Science in the University of Oxford, who expresses it thus: "Resolution of any stress-system into uniform tension and shearing stress.* We have shown that by a suitable change of axes any state of stress can be represented by their principal stresses p_1, p_2, p_3. If now we write

$$\left.\begin{aligned} p_1 &= \tfrac{1}{3}(p_1+p_2+p_3)+\tfrac{1}{3}(p_1-p_2)+\tfrac{1}{3}(p_1-p_3) \\ p_2 &= \tfrac{1}{3}(p_1+p_2+p_3)+\tfrac{1}{3}(p_2-p_3)+\tfrac{1}{3}(p_2-p_1) \\ p_3 &= \tfrac{1}{3}(p_1+p_2+p_3)+\tfrac{1}{3}(p_3-p_1)+\tfrac{1}{3}(p_3-p_2) \end{aligned}\right\} \quad \ldots\ldots(2)$$

we see that the state of stress may be regarded as a superposition of four distinct stress systems. In the first we have three *equal* principal stresses of magnitude

$$\tfrac{1}{3}(p_1+p_2+p_3)=p \ \ldots \ \text{(say)},$$

where p is the mean tension at the point considered. The other component stress-systems are alike in type. In the first (for example) the principal stresses are

$$\pm\tfrac{1}{3}(p_1-p_2),\ 0 \qquad \ldots\ldots(3)$$

so that the mean tension is zero. From § 283 we see that the maximum intensity of shearing stress is $\frac{1}{3}(p_1-p_2)$ and occurs on planes which lie midway between the planes of non-zero principal stress. It is easy to show that the normal component of stress is zero on such planes; accordingly the system represented by (3) may be looked on as a pure shear of intensity $\frac{1}{3}(p_1-p_2)$ and the other two component systems, similarly, as pure shears of intensities $\frac{1}{3}(p_2-p_3)$ and $\frac{1}{3}(p_3-p_1)$."

It is evident that as (p_1-p_2) is one of the terms of equation (1), the difference between the omission of a possible small effect from the "stress perpendicular to the plane of the shear" or of the smaller shear stresses corresponding to equation (3) might be insignificant. Nevertheless, the suggestion cannot be accepted because the separation of a state of stress into volumetric stress and shear stress requires that the shear stresses should be those of a system; i.e. if an elementary cube be imagined at the point of stress, such shears act along its faces parallel to its edges.

Consider the simple case of a material subjected to a tension p, and a simultaneous hydrostatic pressure $\frac{1}{2}p$ in all perpendicular directions as indicated in figure 1. (This is a simple tensile stress, $\frac{3}{2}p$, from which the average volumetric stress has been removed for purposes of simplicity in treatment.) According to the above method of resolution, the stress system is equivalent to three shear stresses $\frac{1}{3}(p+\frac{1}{2}p)$; $\frac{1}{3}(p+\frac{1}{2}p)$; 0, that is $\frac{1}{2}p$ and $\frac{1}{2}p$ on planes inclined at 45° to the axis of p and zero on planes passing through that axis, as indicated in figure 1. It is clear that these shear stresses are not on the faces of one elementary cube and so do not form a system.

For the same set of principal stresses there will be on a plane *ABDC*, figure 2, inclined at an angle α equal to $\cos^{-1}(1/\sqrt{3})$ to p, normal stress t and a shear stress q of values given by

$$t=p\sin^2\alpha-\tfrac{1}{2}p\cos^2\alpha=0,$$

$$q=(p+\tfrac{1}{2}p)\sin\alpha\cos\alpha=p/\sqrt{2}.$$

---

* § 284, p. 274.

If on *ABDC* a square *EFGH* with its sides at 45° to *AB* be taken, the shear stress $p/\sqrt{2}$ is equivalent to a pair of shear stresses $p/2$ acting parallel to the sides of the square as indicated. The edges of three squares thus inclined, oriented equally about the line of action of p, coincide, forming a set of three mutually orthogonal edges containing an elementary cube of which *EFGH* is a face, shown at *a* in the figure. On the faces of this cube there is no normal stress, but only shear stresses, each equal to $p/2$, parallel to its edges as indicated in figure 2*a*. This illustrates a shear-stress system. For the principal stresses given, and for any combination of a volumetric stress with them (e.g. for a simple tension), this is the only possible shear-stress system.

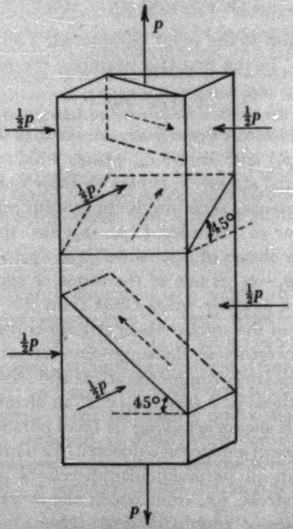


Figure 1.

This equivalent stress system is quite different from that of the type of resolution under consideration. While the stresses of equation (3) are equivalent to shear without normal stress on an element rotated through 45° about p_3, this rotation is inconsistent with similar rotational operations to deal with the other pairs of stresses. A further difficulty also presents itself in this "resolution of any stress system" in that in the state of stress specified by p_1, p_2, p_3 the maximum shear stress is $\frac{1}{2}(p_1-p_3)$, while according to this resolution it is $\frac{1}{3}(p_1-p_3)$: and no form of resolution can alter the maximum shear actually existing. The invariants in this resolution also present a difficulty. For these reasons this type of resolution, despite its apparent helpfulness, must be rejected in the statement of a criterion of failure.

To proceed to the consideration of true shear-stress systems (as exemplified in figure 2*a*), the discussion will be perfectly general if the volumetric stress be taken as zero, so that $p_1+p_2+p_3=0$, when the principal stresses may be written, p, $-kp$,

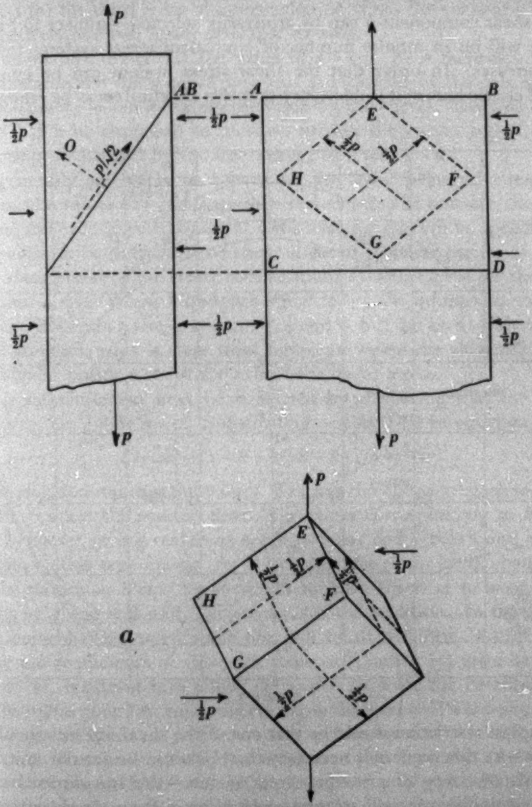


Figure 2.

and $-(1-k)p$, where k will be a fraction less than $\frac{1}{2}$ (since values of k above $\frac{1}{2}$ correspond to former values of $1-k$) and may be taken as positive without loss of generality. Conveniently, and still without loss of generality, the stresses may be taken as 1, $-k$ and $-(1-k)$. If the equivalent shear-stress system be f, g, h, the equating of the invariants of the stress quadrics gives

$$f^2+g^2+h^2=1-k+k^2,$$

and $$2fgh = k(1-k),$$

so that and $$\left.\begin{aligned} f+g &= \sqrt{\{1-k+k^2-h^2+k(1-k)/h\}}\ \ldots, \\ f-g &= \sqrt{\{1-k+k^2-h^2-k(1-k)/h\}}\ \ldots. \end{aligned}\right\} \quad \ldots\ldots(4)$$

As the shear component h can be arbitrarily selected—subject to f and g being real—there will be an infinite number of equivalent stress systems for any set of principal stresses. In order that the shear-stress system can be employed in a criterion of elastic failure it is necessary that this indefiniteness be removed in some manner.

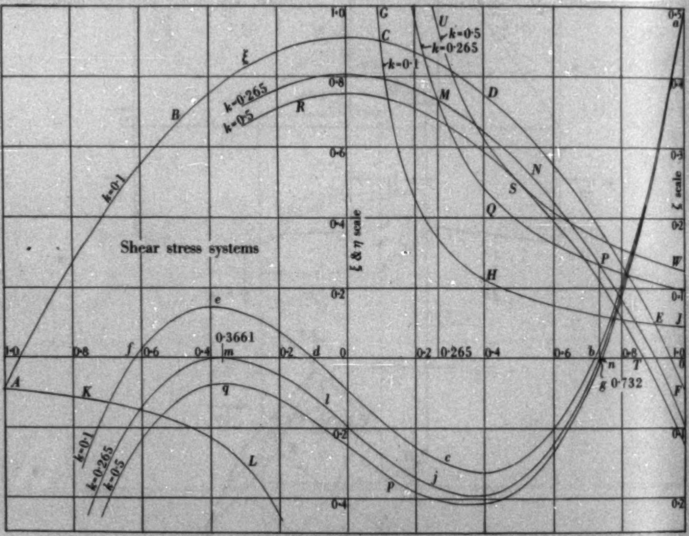


Figure 3.

The simplest restriction would be that one of the shear stresses should be zero. As $fgh = k(1-k)$, this renders it necessary that k should be zero or unity. There is therefore only one type of principal stress system—the two-dimensional case of two equal and unlike principal stresses—which has a shear-stress system having a zero component and this restriction is therefore inapplicable.

For f and g to be real, two restrictions as to the value of h are necessary: ξ, equal to $1-k+k^2-h^2$ must be positive and also it must be greater than η, equal to $k(1-k)/h$, which, by symmetry, need only be regarded as positive.

Let curves corresponding to these equations be drawn, the values of h being taken as abscissae and the ordinates ξ and η perpendicular to the h axis. In figure 3, *ABCDEF* and *GCHEJKL* are a pair of such curves and for real shear-stress systems h must be selected so that the curve ξ is there above the axis Oh and also above the curve η. Thus the limits are given by the points C and E in which the curves intersect. Equating ξ and η gives k and $1-k$ as the abscissae of C and E, and the value -1 for the third point of intersection, A. Thus the value of h is limited to the range between the two lesser principal stresses.

These curves are drawn for the value of $k=0\cdot1$. As k increases, the ξ curve moves downwards and the η curve outwards as indicated by the pair MNP, MQP for which $k=0\cdot265$. When k reaches the value $0\cdot5$ the curves, shown as RST, USW touch at S. If k be supposed to increase further the curves repeat, the former positions being unchanged by the interchange of k and $1-k$.

Thus generally any value of h between the lesser principal stresses may be selected, and the values of f and g are then real and also lie between these principal stresses, and for each selected value of h the axes for fgh will be specifically orientated. It is also to be noted that no shear stress of a system can be as great as the maximum shear stress of the stress combination, $1-k/2$, except for the pure shear case discussed above. When k becomes equal to $0\cdot5$ the infinite number of shear-stress systems becomes a single system, $\frac{1}{2}$, $\frac{1}{2}$ and $\frac{1}{2}$, but the axes will have an infinite number of orientations as is clear from figure 2*a* where the elementary cube may have any angular position about the axis of the major stress.

Let this indefiniteness next be restricted by supposing that the sum of the shears be zero, i.e. $f+g+h=0$. Substituting $-h$ for $f+g$ in equation (4) gives

$$\zeta=2h^3-(1-k+k^2)\,h-k\,(1-k)=0 \qquad \text{......(5)}$$

as the cubic for determining the values. Tracing the curves corresponding to ζ on the diagram renders the matter clear. The curve corresponding to $k=0\cdot1$ is represented by *abcdef*, giving real roots $0\cdot718$, $-0\cdot101$ and $-0\cdot618$, and a true shear-stress system of this type exists. As k increases the cubic curve changes, passing through the particular shape *agilm* which touches Oh at m to *anpq*, for $k=0\cdot5$, which gives only one real root, $On=0\cdot745$, for the equation. In the limiting case *agjlm*, the cubic having equal roots Om will be of the form $(h+x)^2\,(h-2x)=0$. Comparing the coefficients of this with those of equation (5) gives $6x^2=1-k+k^2$ and $4x^3=k-k^2$, resulting in a cubic $4x^3+6x^2-1=0$ for the determination of x. This has the root $0\cdot3661$ for which the corresponding value of k is $0\cdot265$. For values of k exceeding this, there are no shear-stress systems in which the sum of the shear stresses is zero. The curve *agjlm* is drawn for the value $k=0\cdot265$, as also the corresponding pair of curves MNP and MQP, the abscissae of g and P being the same.

As the combinations of principal stresses for shear-stress systems in which $f+g+h=0$ are thus limited, this condition cannot form part of a criterion of elastic failure.

A restriction which would be applicable in all cases is the selection of the largest (or smallest) shear stress available, $1-k$ or k, as one of the system. In the former, the values of $1-k+k^2-h^2$ and $k\,(1-k)/h$ being each equal to k, the other stresses of the system are each $\sqrt{(k/2)}$. The shear stresses associated with k would similarly be each $\sqrt{\{(1-k)/2\}}$.

The change of aspect of the state of stress is easily visualized in such cases. If

OA, OB and OC be the axes of the principal stresses 1, $-k$ and $-(1-k)$ and OF, OG, OH those of the shear stresses $\sqrt{(k/2)}$, $\sqrt{(k/2)}$ and $(1-k)$, the direction cosines of the former relative to the latter are obtained by substituting in sequence 1, $-k$ and $-(1-k)$ for p in the equations

$$lp=(1-k)\,m+\sqrt{(k/2)}\,n,$$
$$mp=(1-k)\,l+\sqrt{(k/2)}\,n,$$
$$np=\sqrt{(k/2)}\,(l+m),$$

giving the resulting scheme of relationship shown in the table.

	F	G	H
A	$1/\sqrt{\{2(1+k)\}}$	$1/\sqrt{\{2(1+k)\}}$	$\sqrt{\{k/(1+k)\}}$
B	$\sqrt{\{k/2(1+k)\}}$	$\sqrt{\{k/2(1+k)\}}$	$1/\sqrt{(1+k)}$
C	$1/\sqrt{2}$	$-1/\sqrt{2}$	0

The first system gives an elementary cube with edges along OA, OB, OC, having normal stresses 1, $-k$, $-(1-k)$ only on its faces. If this cube be supposed to turn round OC, figure 4, through an angle AOH equal to $\cos^{-1}\sqrt{\{k/(1+k)\}}$ so that its edge lies along OH, the position will be that shown by the (broken line) cube LMN, and it will have on the face LM shear stress only of amount $\sqrt{k}$ and upon face MN a normal stress also of amount $1-k$, while that on face LM remains $-(1-k)$. This cube now may be imagined to rotate round OH through 45° into its final position shown by $L'M'N'$, its edges coinciding with OF, OG and OH, when the two normal stresses $\pm(1-k)$ give a shear stress of the same amount and the shear stress $\sqrt{k}$ gives two equal shears $\sqrt{(k/2)}$ parallel with the edges of the cube, upon the faces of which there is no normal stress. The equivalence of the views or resolutions of the state of stress is thus easily visualized.

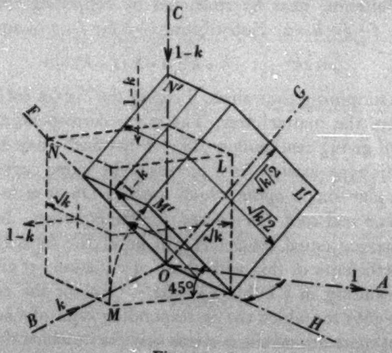


Figure 4.

The case in which the stress $-k$, the point C in figure 3, is selected instead of

$1-k$ corresponds, the equal shears being now $\sqrt{\{(1-k)/2\}}$. Thus the shear stresses of the system under the restriction are the two lesser principal stresses and a fraction $(1/\sqrt{2})$ of the geometric mean between each of them and the greatest principal stress—all being taken to be after the removal of the average volumetric stress.

According to the suggested amendment the criterion would become $q+\mu v=A$, where v is the average volumetric stress, q a stress of the shear system, and A a constant. Comparison with the experimental results given in my paper shows that under the above restriction only the selection of the stress $\sqrt{\{(1-k)/2\}}$ can be made to give any agreement, and that stress would necessitate a minimum value of μ, which, although only about 0·04, would be unexpected. General practical results show, however, that such a criterion cannot be generally true.

The four cases of single and double tension, and single and double compression (i.e. $p, 0, 0 : p, p, 0 : 0, 0, -p$: and $0, -p, -p$) have shear-stress systems of the unique type, so that no ambiguity can occur in their comparison. The shear stress of the system is $\frac{1}{3}p$ in each case, the differences being in the volumetric stress. If the respective yield stresses are f_0, f_{00}, p_0 and p_{00}, the equation gives

$$f_0=3A/(1+\mu),$$

$$p_0=3A/(1-\mu)=\frac{1+\mu}{1-\mu}f_0,$$

$$p_{00}=3A/(1-2\mu)=\frac{1-\mu}{1-2\mu}p_0.$$

As p_{00} cannot be infinite, the maximum value of μ is $\frac{1}{2}$. This gives the corresponding maximum value of the ratio p_0/f_0 to be three. Tests give the value of this ratio for cast iron to vary from four to six. Thus the equation cannot be applicable to metals of the cast iron type.

Thus the primary objection to the suggestion that the average volumetric stress and some constituent of the residual shear-stress system should replace the selected terms in the expression of the criterion is that the shear-stress system is indefinite and that any restriction, such as those considered, made to reduce this indefiniteness would be an artificiality which would require some special justification.

There are then the further objections that no restriction appears satisfactory and finally that a criterion so expressed would be of limited application, all metals having a ratio of compressive to tensile yield stresses greater than three being excluded from it.

For those, and other reasons, the average volumetric stress with the allied shear stresses only, and the corresponding strain representation, were discarded in determining the expression for the results of this research.

As the criterion stands, it avoids objections such as those discussed above and is applicable to all isotropic materials, and all subsequent research appears to substantiate its truth.

REFERENCE

(1) *Phil. Mag.* (July 1900); *Proc. Phys. Soc.* 17 (Sept. 1900).

A MACHINE FOR THE RAPID SUMMATION OF FOURIER SERIES

By C. A. BEEVERS, D.Sc., F.Inst.P.
Dewar Fellow, University of Edinburgh

Communicated by Prof. W. L. Bragg, F.R.S., 16 *January* 1939. *Read* 12 *May* 1939

ABSTRACT. The rapid summation of Fourier series is becoming a more and more urgent problem in x-ray crystallography. A machine is proposed in which the fundamental sine curve of any amplitude may be represented by a number of sets of impulses. A switch can arrange these sets of impulses to correspond to a wave of any wave-number, and the sets of impulses then go into counters. Successive operations of the generator and arranging-switch result in the addition of Fourier terms, the terms being added for all values of the variable at the same time.

§1. INTRODUCTION

The addition of a Fourier series of cosine or sine waves of different wave-numbers and amplitudes has always been of interest in applied mathematics. Also in modern x-ray work on the structure of crystals a frequent process is a Fourier summation of the F-values (which are related to the intensities of reflection) over an area, and the new series invented by Patterson(1) has still further extended the scope for such summations. Also with Weissenberg methods of photography in crystal structure it is not very tedious to obtain rough values of all the general (*hkl*) intensities of reflection. Thus three-dimensional Fourier summations are being increasingly employed in order to utilize all the observations made. Now two-dimensional and three-dimensional summations are most conveniently made by splitting them into one-dimensional summations, but a very large number of these are involved. It is therefore of importance to develop the fastest possible method of making the one-dimensional summation, and to render this method available to all crystal-structure laboratories.

The most speedy method which is generally available at the moment is to use the set of printed strips prepared by Mr Lipson and the author(2), provided that the scope of the strips (range of wave-number and amplitude, and division of interval) is sufficient for the problem on hand. This is usually the case in crystal-structure problems. The strips carry numbers representing all the sine or cosine waves required, and the process consists merely in the selection of the appropriate strips, and the addition of the columns of numbers thus obtained.

Now the selection and addition of strips of numbers is precisely the process which is accomplished in systems such as the Hollerith series of calculating machines. Here the numbers are represented by holes punched in cards, and the required cards can be selected and run through the adding machine at a rate of about one per second. The machine generates sets of electrical impulses from data provided by the cards and prints totals. The writer had an opportunity (provided by Dr L. J. Comrie, of Scientific Computing Service Ltd.) of a thorough investigation of the application of the Hollerith system to Fourier summations. Unfortunately the cost of the adding machine provides a serious obstacle. A central bureau, to which the data would be sent for computation, would be required. Such a system, however, would not be as convenient as computation by the present system (the Fourier strips) except for very long pieces of work. The simplest synthesis would require several days for the results to come to hand, and it is a great disadvantage for the computation to be outside the direct control of the investigator.

It seems however that it would not be very expensive to produce a small machine using the same principle as the Hollerith which would do the required summation with great speed, if full advantage were taken of the relations between the various strips of numbers. Such a machine would generate directly the impulses corresponding to the numbers on any one of the Fourier strips, and pass these into counters. Then it would generate impulses corresponding to the next required strip, and pass these into the same counters, and so on until all the terms have been added. The machine would work without the intermediary of punched cards.

§2. THE GENERATION OF THE FOURIER STRIPS AS SETS OF ELECTRICAL IMPULSES

Consider, first, all the strips having the same amplitude, say the amplitude 45. The 45 sin 1 strip gives the value of $45 \sin \theta$ for values of θ in steps of 6° from 0° to 90°. The number following the word "sine" in the description of the strip is the wave-number, and a strip with wave number 2 and amplitude 45 would give the values of $45 \sin 2\theta$ in 6° steps of θ. It follows that *all* the strips of amplitude 45 contain only numbers which are on the strip of wave-number 1, but the numbers are re-arranged, and may be negative. Some of the strips of amplitude 45 are shown in table 1. If therefore the 45 sin 1 strip is generated as a set of electrical impulses which can be fed into 16 counters (for the different values of θ), any one of the strips of amplitude 45 could be added in if a switch were provided to connect the counters to the impulses in the right order. In the set of strips published by Mr Lipson and the author there are 41 strips of any one amplitude (21 wave-numbers of cosine, and 20 of sine). In order to permit the registration of the negative numbers we may have counters with negative drive magnets as well as positive ones, and connect to the negative sides as required. The connexions may be made by a 41-way switch. Each way or setting of the switch corresponds to one wave-number of cosine or sine, and there are 32 poles, each connected to the positive or negative side of a counter.

We have now to generate only the sin 1 strips of all amplitudes. The beginning and ending of the complete set of sin 1 strips are shown in table 2, each strip being

Table 1

45 S 1	0	5	9	14	18	22	26	30	33	36	39	41	43	44	45	45
45 S 2	0	9	18	26	33	39	43	45	45	43	39	33	26	18	9	0
45 S 3	0	14	26	36	43	45	43	36	26	14	0	$\overline{14}$	$\overline{26}$	$\overline{36}$	$\overline{43}$	$\overline{45}$
……	……	……	……	……	……	……	……	……	……	……	……	……	……	……	……	……
……	……	……	……	……	……	……	……	……	……	……	……	……	……	……	……	……
45 S 20	0	39	$\overline{39}$	0	39	$\overline{39}$	0	39	$\overline{39}$	0	39	$\overline{39}$	0	39	$\overline{39}$	0
45 C 0	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45
45 C 1	45	45	44	43	41	39	36	33	30	26	22	18	14	9	5	0
45 C 2	45	44	41	36	30	22	14	5	$\overline{5}$	$\overline{14}$	$\overline{22}$	$\overline{30}$	$\overline{36}$	$\overline{41}$	$\overline{44}$	$\overline{45}$
45 C 3	45	43	36	26	14	0	$\overline{14}$	$\overline{26}$	$\overline{36}$	$\overline{43}$	$\overline{45}$	$\overline{43}$	$\overline{36}$	$\overline{26}$	$\overline{14}$	0
……	……	……	……	……	……	……	……	……	……	……	……	……	……	……	……	……
……	……	……	……	……	……	……	……	……	……	……	……	……	……	……	……	……
45 C 20	45	$\overline{22}$	$\overline{22}$	45	$\overline{22}$	$\overline{22}$	45	$\overline{22}$	$\overline{22}$	45	$\overline{22}$	$\overline{22}$	45	$\overline{22}$	$\overline{22}$	45

one row of the table. There are 15 columns of numbers in each strip, and each column as we proceed step by step downwards either increases by unity at each step or remains unchanged. The number in each column may therefore be generated by one pole of a switch which moves down 99 contacts. Wherever there is an increase in the number in table 2 the corresponding contact must be live, and wherever there is no increase the contact must be dead. Live contacts are shown by crosses in table 2. With this arrangement, when the switch moves down to a row determined by the amplitude required, the 15 poles generate the required strip as a set of impulses. These impulses then pass along to the wave-number switch, and are there arranged to form a wave of the required wave-number.

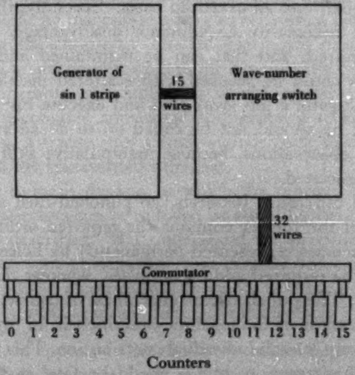
Table 2. sin 1 strips

1 S 1	0	0	0	0	0	0	$\overset{\times}{1}$	$\overset{\times}{1}$	$\overset{\times}{1}$	$\overset{\times}{1}$	$\overset{\times}{1}$	$\overset{\times}{1}$	$\overset{\times}{1}$	$\overset{\times}{1}$	$\overset{\times}{1}$	$\overset{\times}{1}$
2 S 1	0	0	0	$\overset{\times}{1}$	$\overset{\times}{1}$	$\overset{\times}{1}$	1	1	1	$\overset{\times}{2}$	$\overset{\times}{2}$	$\overset{\times}{2}$	$\overset{\times}{2}$	$\overset{\times}{2}$	$\overset{\times}{2}$	$\overset{\times}{2}$
3 S 1	0	0	$\overset{\times}{1}$	1	1	1	$\overset{\times}{2}$	$\overset{\times}{2}$	$\overset{\times}{2}$	2	$\overset{\times}{3}$	$\overset{\times}{3}$	$\overset{\times}{3}$	$\overset{\times}{3}$	$\overset{\times}{3}$	$\overset{\times}{3}$
4 S 1	0	0	1	1	$\overset{\times}{2}$	$\overset{\times}{2}$	2	$\overset{\times}{3}$	$\overset{\times}{3}$	$\overset{\times}{3}$	3	$\overset{\times}{4}$	$\overset{\times}{4}$	$\overset{\times}{4}$	$\overset{\times}{4}$	$\overset{\times}{4}$
……	……	……	……	……	……	……	……	……	……	……	……	……	……	……	……	……
……	……	……	……	……	……	……	……	……	……	……	……	……	……	……	……	……
96 S 1	0	10	20	$\overset{\times}{30}$	39	$\overset{\times}{48}$	56	64	71	$\overset{\times}{78}$	$\overset{\times}{83}$	$\overset{\times}{88}$	$\overset{\times}{91}$	$\overset{\times}{94}$	$\overset{\times}{95}$	$\overset{\times}{96}$
97 S 1	0	10	20	30	39	48	$\overset{\times}{57}$	$\overset{\times}{65}$	$\overset{\times}{72}$	78	$\overset{\times}{84}$	$\overset{\times}{89}$	$\overset{\times}{92}$	$\overset{\times}{95}$	$\overset{\times}{96}$	$\overset{\times}{97}$
98 S 1	0	10	20	30	$\overset{\times}{40}$	$\overset{\times}{49}$	$\overset{\times}{58}$	$\overset{\times}{66}$	$\overset{\times}{73}$	$\overset{\times}{79}$	$\overset{\times}{85}$	$\overset{\times}{90}$	$\overset{\times}{93}$	$\overset{\times}{96}$	$\overset{\times}{97}$	$\overset{\times}{98}$
99 S 1	0	10	$\overset{\times}{21}$	$\overset{\times}{31}$	40	49	58	66	$\overset{\times}{74}$	$\overset{\times}{80}$	$\overset{\times}{86}$	90	$\overset{\times}{94}$	$\overset{\times}{97}$	$\overset{\times}{98}$	$\overset{\times}{99}$

In order to be able to insert into the counters a wave with a negative amplitude (these waves are printed on the backs of the Fourier strips) the wires from the wave-number switch to the counters must pass to the counters by way of a commutator, which will interchange negative for positive and vice versa. This requires a switch which is the equivalent of sixteen Pohl commutators.

§3. THE DESIGN OF SUITABLE COUNTERS

The 16 counters must be capable of reliable operation at speeds of 50 to 100 counts per second. This is faster than the usual type of electromagnetic counting mechanism, but no faster than the uniselector switches used in automatic telephony. The counters must also have a zeroing device, and must be capable of being operated in both directions. It is quite probable that suitable counters could be made up from uniselector switches, but it would perhaps be more satisfactory in the long run to build special instruments. Since the electromagnetic mechanism has only to turn a counting wheel, these counters could be made faster than uniselector switches, in which a number of switch arms have to be turned. A two-wheel counter with one rotation of the input wheel per 100 impulses would be satisfactory, so that this wheel would give tens and digits. The second wheel, moving on one place for every revolution of the first, would give the thousands and hundreds. Such a counter would be a simple modification of the Hope-Jones electrical-impulse dial movement, which has already been shown to work well at 100 impulses per second(3).



The figure gives a diagram of the whole machine, and it is considered possible to make such a machine which will add a ten-strip summation within a time of 10 sec., leaving the operator only to copy down the answer. At a greater cost the machine could be made to print the totals, and its range could be extended in any way.

REFERENCES

(1) Patterson. *Z. Kristallogr*, **90**, 517 (1935).
(2) Lipson and Beevers. *Proc. Phys. Soc.* **48**, 772 (1936).
(3) Hope-Jones. *Electric Clocks*, p. 125 (N.A.G. Press, Ltd.).

DISCUSSION

Dr L. J. Comrie. I have always advocated, as a policy, that the capabilities of existing commercial machines should be fully exploited before the laborious and costly task of designing and making special apparatus is embarked upon. In this case the nature of the problem points to the use of punched-card machines, which have been successfully applied to Fourier synthesis in the summation of the harmonic terms in the moon's motion. I am not convinced by the curt dismissal of these machines. First, the cost of using them is not likely to exceed the cost of designing and constructing the machine described, if due allowance be made for the time of the designer, and all overhead costs. Secondly, the centralizing of the computations is not necessarily an ultimate disadvantage. It means that one machine only is required, and that can be maintained and run with a higher efficiency factor than could be secured with scattered installations working part time only. The investigator's job is to plan and execute his observations, and to interpret the results; he should not be called on to do an amount of arithmetic comparable with his observations, because manipulative skill and numerical skill are not necessarily associated.

The fundamental difference between the punched-card system and the proposed machine is that the former contains the required ordinates of waves of all amplitudes and frequencies represented permanently by holes in cards, which can be used to generate the required adding impulses, whereas the latter generates its own impulses ad hoc. The punched-card machines, of course, print their results.

The principal weakness of the machine described is, in my view, that it yields two-figure accuracy only in each individual contribution. This would not be serious in a one-dimensional summation, but it would be fatal in two-dimensional and three-dimensional summations. It is true that the amplitudes of the waves which we are compounding are known to two significant figures only. But in two-dimensional summations we may easily generate ten or more new amplitudes from the given data by adding ten rounded-off numbers; each of these then becomes the argument for a further set of waves, so that the final summation may easily contain an effect equivalent to summing a hundred rounded-off numbers. In other words, even the first figure ceases to be reliable. With three-dimensional summations the derived figures would be almost meaningless. It is necessary, therefore, to annex ciphers to the original amplitudes, carry a third figure throughout, and round off the last figure at the end; this indeed is done by certain investigators. To work with three figures in the proposed machine it would be necessary to generate numbers by one pole of a switch that moved down 999 contacts, and the wheels of the counters would need 1000 teeth each.

From the final summations, contours of equal values are interpolated. It seems unsound, therefore, to include frequencies as high as 20 when the interval is 6°, because that implies that we can interpolate a sine wave from points 120° apart. The theoretical limit is one radian, but in practice, to secure a reasonable convergence in the differences, the limit is between 20° and 30°. Thus we require several times as many counters as have been proposed.

If the machine generates only a quarter-wave, there will be many occasions when the sums and differences of written totals have to be formed and written by hand. As the function of the machine is to reduce human labour, it would seem better to carry the waves to 180° if possible, as is done in the punched-card system.

In conclusion, I am surprised that crystallographers do not prefer an existing machine now to waiting several years for a machine that is still in the early stages of design.

Dr C. E. WYNN-WILLIAMS. The author has shown by means of a model that the proposed scheme is practicable. An objection which might be raised, however, is that the counting might be inaccurate owing to occasional faulty operation of the generating switches or the counters. The trouble might possibly be avoided by adopting some form of guard circuit, such as for example the type of circuit used in certain telecontrol apparatus, although this would render the apparatus more complicated. In the telecontrol apparatus in question a receiving switch remote from, and operated by, a transmitter switch signals back to the latter that it has responded correctly to the transmitted impulses.

A point of interest to me is the way in which the number of contacts required on the generating uniselectors can be reduced by adopting the circuit arrangements which the author has described. If I understand the proposed scheme properly, six 25-contact banks on four uniselectors, that is, only 600 contacts, are sufficient to enable one to reproduce, electrically, the whole of table 2, which contains, as will be seen, 99 × 16, or 1584 separate pieces of information.

Dr I. FANKUCHEN. It appears to me that Beevers and Lipson's contributions to the computation of Fourier series have not been properly appreciated. In the period prior to their work, the computation of a double Fourier series was a nerve-racking month-long job, whereas the use of the strips, in the hands of an experienced user, shortened the period to about one day. Dr Comrie's two leading objections to the strip method are certainly not valid. The author has refuted one by showing that the errors introduced by rounding off the numbers to the nearest unit are really negligible and not, as Dr Comrie has suggested, of any great importance. Dr Comrie's other criticism, to the effect that taking the unit of subdivision every 6° is not sufficient for the components of higher frequency, is based on a misconception of what is wanted of the strips. It is obviously true that a high-frequency component cannot be adequately represented by a 6° interval, but the strips are not intended to be used to draw out the components. They are merely intended to represent the value of the component at a certain number of prefixed points. The necessary number of points is determined, as Beevers and Lipson have shown in their papers, by the size of the unit cell to be studied, and for all reasonable unit cells 60 divisions per side of cell is certainly adequate. Thus, for a cell 20 A. by 20 A. there would be 3600 points, and an average atom 3 A. in diameter would be represented by about 60 points. Of course for the larger unit cells found in protein crystals, a greater number of divisions would be desirable. It must be emphasized that the frequency of the components cannot, as Dr Comrie has suggested, influence the choice of the number of divisions.

The increasing use of Fourier methods as a means of testing trial structures and as a method of successive approximations makes it desirable that any devices used for the rapid summation of such series should be of a nature which will make possible their use in individual laboratories. A central computing bureau for work of high accuracy, for large unit cells or other special cases, may be desirable, but the real need is for a small comparatively cheap unit for the individual laboratory.

Dr E. G. Cox. In connexion with the x-ray analysis of crystalline organic compounds I have now under construction a machine designed primarily for the calculation of structure factors, i.e., for the evaluation of expressions of the type

$$F_{hl}=\sum_{x}\sum_{z} f_{xz} \cos 2\pi\left(\frac{hx}{a}+\frac{lz}{c}\right). \qquad \text{......(i)}$$

A full description will be published shortly; briefly, the quantity f is represented by a capacity, and the cosine term by an e.m.f. which is caused to vary harmonically by means of a suitable potentiometer. By the application of this e.m.f. a charge proportional to the product of f and the appropriate cosine term is communicated to each condenser, and the summation is simply effected by measuring the total charge on all the condensers.

From the form of equation (i) it can be seen that the machine is equally applicable to the summation of Fourier series; its accuracy is not likely to be sufficient for the final analysis of a crystal structure, but will be adequate for the numerous preliminary syntheses which are necessary. Its advantages are that:

(*a*) A summation of the type (i) can be made in a few seconds, whatever the number of terms.

(*b*) The summations can be calculated for any increments in the variables, however small. A machine such as the author's, employing fixed intervals in the argument, cannot be used for structure-factor calculations.

(*c*) If necessary, syntheses can be carried out for a small region only of the unit cell, e.g. in a region in which one or more atoms is expected.

AUTHOR'S REPLY. Dr Comrie has stated that the proposed machine would be inaccurate for two-dimensional summations, and even says that in three-dimensional summations the results would be almost meaningless. This is not the case. It is of course true that the absolute errors increase with the number of strips used, but the summation totals also increase over regions where there are atoms, and these are the important regions. In fact the percentage accuracy increases with the number of dimensions. Since the data to be inserted into the synthesis are known only to two figures (and even this refers only to the most careful spectrometric observations), there can be no question of any rounding-off error at all, for there has been no rounding-off. We have repeated Robertson's synthesis on benzoquinone using two figures, and compared it with Robertson's results, for which three figures were used. There was no significant difference whatever. Any Fourier synthesis made with the Fourier two-figure strips is accurate enough to show false peaks which are due to errors in the data and not in the computation. While this is the case there is no point in a three-figure accuracy.

The incorporation of frequencies as high as 20 in the proposed design of machine is also criticized. It is true that interpolation on a sine curve defined by points 120° apart is very inaccurate, but the interpolation is in fact carried out on the sum of the series, and this sum represents a curve which may be varying comparatively slowly although it possesses high-frequency components. The only necessity is that the curve should be evaluated at sufficient points to define the atoms. This has already been considered in the literature, and the proposed machine would be inadequate for a final synthesis when the cell edge is greater than about 15 A. As is mentioned in the paper, however, its range can be extended, and it is thought that such an extension of range could be carried out very cheaply.

Dr E. G. Cox's machine is an interesting one, and will no doubt be very useful for structure-factor evaluations, but it seems unlikely that it can be made very fast for Fourier synthesis. To adjust this machine for each (xz) value, operate it, and take the total charge on all condensers would probably occupy at least 20 sec., and to do this 1800 times, for each (xz) value over a half unit cell, would mean a total time of 10 hr. This is comparable with the time taken now with the Fourier strips. The principle could perhaps be applied to a one-dimensional machine to give faster Fourier synthesis, but a better electrical system would probably be to produce Fourier waves by devices similar to those used in electric organs and to add them on an oscillograph screen.

Note added later. I now feel that there is work for machines of two types: (1) a small machine in each crystallographic laboratory capable of working through a quarter-wave, with unit angle 6°, frequencies up to 20 and amplitudes up to 99; and (2) a more elaborate central machine, such as the Hollerith, capable of working through a half-wave, with unit angles of 3° or less, frequencies as high as are required by complex structures, and amplitudes up to 999.

MEASUREMENTS OF THE INTENSITY-DISTRIBUTION OF THE WHITE X RADIATION REFLECTED FROM A CRYSTAL; WITH A NOTE ON THE DISPERSION OF THE ATOMIC SCATTERING FACTOR OF ZINC NEAR THE *K* ABSORPTION EDGE

By J. C. M. BRENTANO, JOAN HONEYBURNE
and J. K. BERRY

Received 16 *December* 1938. *Read in title* 28 *April* 1939

ABSTRACT. The reflection of the radiation emitted from an x-ray tube by a crystal in an x-ray goniometer gives origin to a pattern of lines standing on a general background. This background is constituted by white radiation emitted by the tube and regularly reflected from the crystal, and by radiation of different origins which has not been subjected to spectroscopic separation. The paper describes a method by which these two constituents of the background can be assessed separately, so that an accurate measurement of the intensity-distribution of the reflected white radiation becomes possible. This method is applied to examine the discontinuities of the background which one of us in collaboration with A. Baxter has indicated as a source of error in the evaluation of the intensities of reflection lines. Microphotometer records illustrating the conditions discussed in a previous paper are reproduced. The method is further discussed in its application to the determination of the dispersive change of the atomic scattering factor in the vicinity of an absorption edge; by its use *f* values for wave-lengths very close to the edge can be obtained. The results of experiments on the long-wave side of the zinc *K* absorption edge are compared with the wave-mechanical dispersion curve.

§ 1. INTRODUCTION

In a previous paper by one of us in collaboration with A. Baxter⁽¹⁾, attention has been drawn to the discontinuities in the white x radiation reflected from a crystal and to the effect of such discontinuities on the evaluation of the intensities of x-ray reflections. It was then shown that considerable errors occur when these discontinuities are not taken into consideration. Such errors arise when reflection lines fall near to discontinuities of the white radiation which is recorded as background either on a photographic film or with an ionization chamber. While in many instances this source of error can be eliminated by avoiding reflections so affected, or by compensating the discontinuities by means of appropriate filters, in other instances the best course is to measure the background steps and to allow for them in the evaluation of the reflection lines.

The present paper deals with a method for observing and measuring these discontinuities under definite conditions and for obtaining data as to their magnitude. This method gains additional interest in that it may be used in the determination of the dispersion of the atomic scattering power, i.e., of the change of atomic scattering power with wave-length in the neighbourhood of an absorption edge of the scattering atoms. Determinations of this kind are usually carried out by comparing the intensities of reflections obtained for some characteristic wave-lengths which happen to fall sufficiently near to the absorption edge. By making use of white radiation a continuous exploration of the range up to a close proximity to the absorption edge becomes possible.

§ 2. THE WHITE X RADIATION AS RECORDED BY AN X-RAY CRYSTAL GONIOMETER

The discontinuities of the white radiation which we have in view are actually not inherent in it from its origin, but are introduced by selective absorption in the materials through which it passes and by sudden changes in the sensitivity of the recording medium. These changes correspond to the absorption edges of silver and bromine in the case of the photographic emulsion, and to similar edges peculiar to the atoms of the gas contained in an ionization chamber. In a previous paper the origin of these various discontinuities has been discussed in detail⁽¹⁾.

We can obtain the spectral distribution of part of this white radiation by reflecting it from the face of a single crystal rotated in the centre of an x-ray goniometer, when the white radiation incident on the crystal will be deflected according to its wave-length λ through an angle 2θ determined by the relation

$$\sin\theta = n\lambda/2d, \qquad \text{......(1)}$$

where d is the spacing of the lattice planes represented by the reflecting surface and n is an integer which determines the order of the reflection. The operation of such an arrangement differs, however, from that of a spectrograph used in the visible range, in that the reflecting crystal is apt to add to the number of discontinuities observed by introducing discontinuities of its own, which may actually form the object of a special investigation, and further because only part of the radiation is regularly scattered or reflected by the crystal through definite angles in conformity with equation (1), while another part is scattered without being subjected to spectroscopic separation. In the latter part are comprised the radiation which leaves the crystal with changed wave-length as Compton scattering; the fluorescent radiation emitted by the constituents of the crystal when the incident radiation is of a wave-length sufficiently short to excite any characteristic radiations in the atoms of which the crystal is built; the general scattering associated with the heat motion of the crystal; and any other radiation scattered under such conditions that the crystal does not act as a regular grating.

This radiation of various origins, which for the purpose of simplifying the present discussion we may designate by the name of "irregularly scattered radiation", is not subjected to a spectroscopic separation by the crystal, and any wave-length appearing in it is distributed as a smooth function of angle over an extended range. Such irregularly scattered radiation is superposed on the regularly scattered or reflected radiation and thus affects any measurements of the distribution of intensity with wave-length of the latter.

§ 3. THE SEPARATION OF THE REGULARLY REFLECTED AND OF THE IRREGULARLY SCATTERED RADIATION

We can separate the regularly reflected from the irregularly scattered radiation by recording the latter outside of the definite locus within which the regularly reflected radiation is confined. This is done by placing a cylindrical rotating screen *D* in a photographic x-ray goniometer in front of the recording film *F*, as indicated diagrammatically in figure 1. The screen *D* is provided with a narrow slit cut at a slanting angle ϵ to the axis of rotation, and it is rotated at an angular speed twice the speed of the single crystal which is placed in the centre of the goniometer and reflects the radiation. The screen then prevents the scattered x rays from reaching the photographic film except for the narrow strip uncovered by the slit. With a parallel x-ray beam entering the goniometer in a direction normal to its axis, regular reflection occurs in a plane normal to the axis of the instrument and in a direction satisfying the relation (1). If the screen is so set that for a given orientation of the crystal a point *A* of the slit is aligned with the regularly reflected beam, then this beam will be recorded on the photographic film; and as the rotation of the screen is coupled to that of the crystal so as to conform with the rate of angular displacement of the reflected beam, the regularly reflected radiation will be recorded on the photographic film as a narrow band corresponding to the displacement of *A*. For convenience of the discussion we call this band "the band of regular reflection".

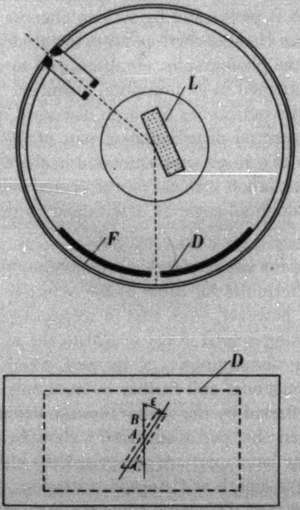


Figure 1.

Points on the photographic film above and below, say *B* or *C* on a line drawn through *A* parallel to the axis and at a distance *h* from the band, are covered by the screen when the slit is in front of *A*. This setting of the slit, as we have assumed, corresponds to a setting of the crystal for which the points *B*, *A* and *C* are situated in the direction of the regularly reflected beam. If now the slit is displaced so as to uncover *B* or *C*, then, owing to the coupling between the rotation of the screen and that of the crystal, the crystal will be turned out of the position of regular reflection through an angle which in angular measure is expressed by $h \tan \epsilon/2s$ when s is the radial distance of the screen from the axis of the instrument. In consequence *B* or *C* will not receive the regularly reflected radiation but only the irregularly scattered radiation corresponding to the new position of the crystal. For any given orientation of the crystal this irregularly scattered radiation is distributed as a smooth function of the angle of scattering; further, for any given angle of scattering the irregularly scattered radiation also is a smooth function of the orientation of the crystal, except for crystal-settings for which the crystal screens off or nearly screens off the scattered radiation from reaching the film, settings which need not be considered here. It can thus be seen that by measuring the intensity of the irregularly scattered radiation as recorded on corresponding points above and below the band of regular reflection, we can interpolate the intensity of irregular scattering for the band itself. In this way we can subtract the irregularly scattered radiation from the intensity observed for the band of regular reflection and so arrive at the intensity-distribution of the regularly scattered radiation.

The actual goniometer used has been described by one of us⁽²⁾. The single crystal took the place of the powder layer, and the rotating screen with the slanting slit was mounted on the rotating platform provided. It is essential that the rotating screen should move with great steadiness and at the correct rate of angular motion. This was secured by mounting the platform on ball bearings and by using the driving mechanism described in the paper⁽²⁾. It was also essential that the width of the slit and its inclination to the axis of rotation should be uniform throughout its length. The slit was milled as a true helix on the correctly centred screen.

§ 4. THE RECORDS TAKEN WITH THE SLANTING SLIT AND THEIR INTERPRETATION

Figure 2 is an example of a record of the radiation from an x-ray tube with copper anticathode reflected from the (0002) face of a zinc crystal. The width of the slanting slit used in this exposure was 4 mm. and the angle ϵ was 45°. The dark band on the figure corresponds to the band of regular reflection. The strong lines are the copper *K* lines, while the fainter lines are due to contamination of the anticathode and belong to the tungsten *L* group. A sudden change of density takes place for the wave-length corresponding to the zinc absorption edge; this is due to the increased absorption in the zinc crystal on the short-wave side. A similar but less pronounced change takes place at the copper *K* edge. In interpreting such records we must consider that the incident beam is not actually parallel but diverges from the entrance slit of the goniometer, and that we have thus the conditions of the

Bragg reflection from an extended single-crystal surface rotated about an axis lying in its surface. This implies that as the crystal is rotated the reflection takes place from various points of the surface in succession. The screen which rotates with the crystal is thus displaced through a finite angle while any particular wave-length is being reflected; this increases the width of the band of regular reflection. A further less definite extension of this band is produced by irregularities of the mosaic crystals forming the crystal surface, irregularities which may consist either in a faulty angular orientation of the crystallites or in a wrong location of the reflecting elements in front of or behind the plane containing the axis of rotation. By limiting the angular aperture of the beam and by choosing suitable crystals this expansion of the band outside the actual height determined by the width of the slit can be reduced. In practice it is found, however, that this extension of regular reflection and its superposition on the irregular scattering outside the band of regular reflection proper is of little inconvenience, and that it can easily be allowed for.

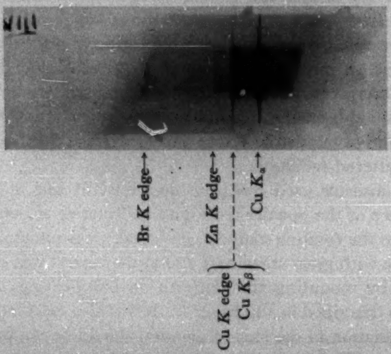


Figure 2. Record of x radiation from a copper anticathode reflected from a zinc crystal, showing two types of density steps in the background of white radiation.

The means for making this correction is found in the reflection lines of characteristic radiation, some of which can be seen in figure 2 to extend beyond the regular reflection band. These lines are due to regular reflection only, so that, by measuring the relative intensities of the reflection lines inside the band and at a certain distance from it, the fraction of the intensity of the regularly reflected radiation which is received at any particular distance outside the band of regular reflection can be found and allowance can be made for it before the intensity of the irregularly scattered radiation is arrived at. This in turn can be subtracted from the intensity of the band in order to arrive at the true distribution of the regularly reflected white radiation.

It will thus be seen that the separation of the irregularly scattered from the regularly reflected radiation can be done by a process which in principle involves successive approximations. Owing to the fact that both the irregularly scattered radiation and the regularly reflected radiation falling outside the band of regular reflection are in general small compared with the intensity within the band, the correction can be carried out with adequate accuracy by a simple operation. The amount of irregularly scattered radiation superposed on the regularly reflected radiation in the central band varies with the width of the slit, and can be reduced by making the slit narrow. The reduction of general incoherently scattered background as compared with exposures taken without a screen is most striking, in particular when conditions are such that excitation of the characteristic radiation of some constituent of the crystal takes place. In such cases the screen can be used as an efficient means for reducing to a minimum the background of reflection patterns obtained with a single crystal.

§ 5. THE DISCONTINUITIES OF THE BACKGROUND OF X-RAY CRYSTAL REFLECTION PATTERNS

With a sufficiently narrow slanting slit the intensity of the various components of the regularly scattered radiation is reduced to such an extent that it need not be allowed for when only a qualitative survey of the discontinuities in the distribution of the regularly reflected white radiation is desired. Figures 3 and 4 are examples of microphotometer records taken at constant height along the band of regular reflection. Figure 3 represents part of a general survey and covers the range of wavelengths from 0·88 A. to 1·540 A., reflected from a zinc crystal. The x-ray tube had a copper anti-cathode contaminated with tungsten; wide collimator slits were used. Two types of density steps can be seen. The step on the extreme left of figure 3 corresponds to the absorption edge of bromine; it is due to the greater absorption in the film with enhanced photographic sensitivity for wave-lengths shorter than the edge. Such steps are characterized by greater density on the short-wave side.

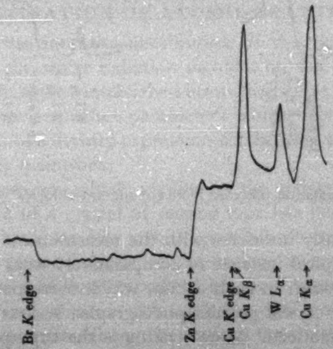


Figure 3. Density steps in radiation reflected from a zinc crystal.

On the right of the figure can be seen the absorption step of zinc, due to selective absorption in the reflecting crystals, and a further step near the copper K_{β_2} line due to selective absorption in the anti-cathode. Such steps show a reduced density on the side of shorter wave-lengths. The significance of these steps and of their study for the evaluation of the intensities of crystal reflection patterns has already been discussed in another paper(1).

As an illustration of what was said there, figure 4 reproduces comparative records of white radiation taken (*a*) with a sodium-chloride crystal and (*b*) with a zinc crystal. Conditions of better definition were here satisfied and two photometer tracings were taken at different heights of the band of regular reflection, so as to provide a check on any accidental irregularities. It will be seen that the copper K absorption edge seriously interferes with the measurement of the copper K_{β_1} and K_{β_2} lines. This is a typical instance of comparatively weak lines falling near to an important absorption step; in such a case, which corresponds closely to the conditions often found in crystal powder photographs, serious errors arise if the discontinuity remains unnoticed. More striking is the superposition of the tungsten L lines and the zinc K absorption edge. The record taken with sodium chloride shows the lines in their relative intensities, while on the record taken with a zinc crystal the zinc absorption step seriously interferes with their evaluation; this is particularly so in regard to the tungsten L_{β_1} line, corresponding to a wave-length of 1·279 A. which falls near to the absorption step of zinc. On the sodium-chloride record this line is shown as the strongest of the four tungsten L lines placed together in a close group, whereas on the zinc record the line is scarcely visible, owing to the proximity of the zinc K absorption edge at 1·2806 A. and to the decrease of the atomic scattering factor in the vicinity of the edge. The technique of eliminating the irregularly scattered radiation by means of the slanting slit opens the way to a

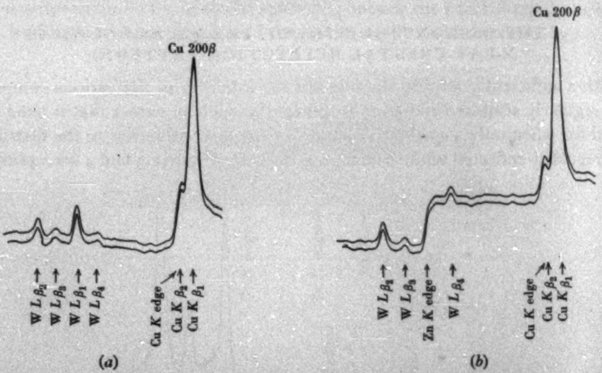


Figure 4. Intensity of radiation reflected from (*a*) sodium-chloride crystal and (*b*) zinc crystal.

closer examination of such conditions. Typical cases are reflections for which wrong values of the intensity are observed owing to the vicinity of absorption steps. By means of comparative records cases of this kind can be brought to a definite analysis and a separate evaluation can be carried out of the reflection line and of the absorption step. To avoid repetition, reference may be made to the previous paper(1) where these conditions are discussed in detail. When quantitative measurements are desired the microphotometer records taken by absorption photometry must be translated by means of a blackening curve into x-ray intensities, and the evaluation must be carried out from the tracings so established. For the qualitative estimation of the conditions represented by the records reproduced in figures 3 and 4 it should be pointed out that when very unequal densities are recorded by microphotographic methods the ordinates are not proportionate to the intensities, but a distorted intensity scale is used which makes it possible to obtain sufficiently exact measurements in that range of densities which is of particular interest. This distortion is automatically provided by absorption photometry and can be obtained for scatter photometry by a special manipulation. Such conditions of distortion existed here and the heights of the strongest lines are therefore not proportionate to the heights of the background steps of the white radiation, although these latter together with the weaker tungsten lines are represented to a good approximation on a linear scale.

§ 6. THE DETERMINATION OF ATOMIC SCATTERING FACTORS

We can compare the intensity-distribution of the regularly reflected white radiation with that of the white radiation incident on the crystal. Between these two quantities a relation can be established into which enter the structure or atomic scattering factor and various other parameters which determine the intensity of reflection. This relation or intensity expression can be used to evaluate the structure factor in terms of other quantities.

The intensity expression which applies to the reflection of monochromatic radiation from the face of a crystal of mosaic type has been discussed by Bragg, Darwin, Ewald and others; we may refer in particular to the discussion given in a paper by Bragg, Darwin and James(3) on the various aspects of the problem. The intensity of the reflection is there expressed by the integrated reflection obtained when the crystal surface is rotated with uniform angular velocity through a sufficiently wide angle to comprise all angular settings for which reflection takes place. The reader is referred to the original paper for a more detailed discussion, and in particular for an account of the conditions under which the intensity of the monochromatic radiation is observed. It will be seen that the results of that discussion can be applied to our case of white radiation. Our conditions differ, however, in that instead of observing the integrated reflection received in an ionization chamber with an entrance slit of unit height and a width sufficient to receive the whole of the monochromatic reflection, we have primarily a source of white radiation and we observe its intensity-distribution after reflection as a photographic density which, when interpreted by means of the blackening curve, represents the quantity of radiation received on a strip of unit height. The reflected radiation is thus measured by a term Pdl or $pd\gamma$, where P and p are respectively the quantities of radiation received in unit time per unit length l of the strip and per unit of the deflection angle γ, while the incident radiation is measured by a term $I\,d\lambda$ to which corresponds an amount $i\,d\lambda$ received as scattered radiation on the strip. A link has to be established between p and i. From equation (1) we have

$$d\theta = n\,d\lambda/2d\cos\theta$$

and thus

$$d\gamma = 2d\theta = \frac{n}{d\cos\theta}\,d\lambda.$$

Introducing the value of d from equation (1), we obtain

$$d\gamma = \frac{2\tan\theta}{\lambda}\,d\lambda. \qquad \text{......(2)}$$

We have thus an intensity expression

$$p_\gamma d\gamma = \frac{\lambda}{2\tan\theta}\,R_1 R_2\,(I_\lambda)\,d\lambda \qquad \text{......(3)}$$

in which $p_\gamma d\gamma$ replaces $i_\lambda d\lambda$. R_1 corresponds closely to the term ρ used in the paper by Bragg, Darwin and James for the integrated reflection, and is thus a function of $N, f, \lambda, \theta, \mu$, and general constants; this term further takes account of heat motion and of the particular experimental conditions. The quantities entering into R_1 are those which have to be considered in any measurements of the intensity of reflection from a single crystal; for a crystal of mosaic type constituted of sufficiently small crystallites the structure factor or atomic scattering factor enters into this expression as f^2. $R_2\,(I_\lambda)$ takes account of the intensity-distribution of the white radiation emitted by the X-ray source; it takes the place of a term measuring the intensity of a monochromatic beam in measurements where not white radiation but characteristic radiations of the anticathode are used.

In its most general aspects the problem of introducing values of $R_2\,(I_\lambda)$ and the application of expression (3) to the determination of atomic scattering factors corresponds to the analogous case in which monochromatic radiations are used, and the difficulties encountered are of the same kind. A considerable simplification, however, occurs when we are primarily concerned not with the determination of the actual scattering factor but with its change with wave-length. It then becomes possible to make use of a crystal constituted of such atoms that the scattering in the region explored is known and regular, for the purpose of establishing an intensity-distribution curve with which the intensity-distribution curve of the radiation reflected from the crystal giving rise to dispersive scattering can be compared; an explicit determination becomes thus unnecessary. One particular point which simplifies evaluations by the method we have described is that for the whole range of wave-lengths the same reflection is observed; thus the factor $(\sin\theta)/\lambda$, and in consequence those terms which are functions of it, in particular the heat factor, are constant for the whole range explored.

§ 7. THE DISPERSION OF THE *F* VALUE FOR ZINC ON THE LONG-WAVE SIDE OF THE *K* ABSORPTION EDGE

In order to explore the possibility of using the method to measure the dispersive change of the atomic scattering factor in the neighbourhood of an absorption edge, we have carried out some measurements on the white radiation reflected from a single zinc crystal.*

Previous determinations of the change of the atomic scattering factor with wave-length were derived from the reflected intensities of characteristic radiations having wave-lengths near to that of the absorption edge. Absorption terms were generally eliminated by using the method of mixed powders and the uncertainty which arises when the results of measurements are referred to explicit *F* values could be avoided by comparing intensities of reflections for different wave-lengths by means of a comparison substance for which the absorption edges are situated far outside the range of wave-lengths concerned. This method has been evolved by one of us in collaboration with Baxter(4) and applied to measurements in the *L* range. It has been adopted by Allison and Jesse(5) and by Jesse(6) in their more recent measurements in the *K* range.

The present position with regard to the results of experimental determinations is that they qualitatively agree with the change of atomic scattering factor in the neighbourhood of an absorption edge indicated by wave mechanics. Experiments confirm the decrease of scattering factor in the immediate neighbourhood of the edge, with a rise on either side to values which for wave-lengths remote from the edge differ by the effective electron number of the particular shell; this is in accordance with wave mechanics(7, 8); it has already been accounted for in the classical treatment of Kallmann and Mark(9). But further, a reasonably good agreement exists between the wave-mechanical dispersive scattering curve and measurements for different wave-lengths near the edge, so far as such measurements are available. Thus little doubt exists that the experimental results and the wave-mechanical treatment of x-ray dispersion agree in the more general features. The finer points of the theory, in particular the effect of the higher terms in Hönl's expressions or, to use Williams's formulation, the differentiation between the various assumptions as to the distribution law of the virtual resonators related to photoelectric absorption, require further measurements for a sufficient number of wave-lengths extending from both sides near to the edge. Such a close range is difficult to obtain with characteristic radiations.† If use can be made, however, of white radiation, a much closer approach to the edge becomes possible, and further, what is actually more important, a continuous sequence of values can be obtained so that experimental data for any part of the dispersive scattering curve are made available.

---

* After this paper had been prepared W. A. Bruce(12) published an investigation on the dispersive scattering effect near the zinc *K* absorption edge. This investigation was based on measurements taken for three wave-lengths 0·44 A., 0·71 A. and 1·54 A., and is thus rather concerned with the verification of the wave-mechanical curve for the dispersive effect in its more general aspects than with the detailed shape of the curve which we consider in our discussion. Bruce finds agreement between the three experimental values and Hönl's curve. This can be taken as an additional indication of the correctness of the wave-mechanical curve in its general features, which is the basis of our treatment.

† It should be noted that, in a general way, for each *K* absorption edge can be found a characteristic radiation situated on the long-wave side very near to it, namely the K_{β_2} radiation of the same element. This is, however, so near to the edge that in measuring it the difficulties are encountered which were discussed in detail by Baxter and by one of us and were referred to also in the earlier part of this paper. The evaluation of the intensity-distribution of a smooth background can in fact be extended to wave-lengths nearer to the edge than the measurement of characteristic lines.

When we apply the method here discussed to examine the dispersion of atomic scattering factors, we must give consideration to the part played by primary and secondary extinction and by the absorption coefficient for the particular wave-lengths used, which quantities all enter in the interpretation of the measured intensities of reflection from an extended crystal. It is difficult to assess these quantities correctly, and it was in fact this particular difficulty which induced one of us(10) to develop the method of mixed powders, whereby the evaluation of the measurements becomes independent of extinction and of absorption effects provided the powders are sufficiently fine. Correct data for extinction and absorption are, however, essential in the first place only then when we require absolute values of scattering factors, or when we make comparative measurements of scattering factors from different substances or of different reflections of unequal strength from one and the same substance. The explicit knowledge of these data is required to a much lesser degree when we are concerned with the comparison of intensities of reflections obtained from the same surface under such conditions that the intensities vary by small amounts only, which is the case here when we are determining the dispersive change of the scattering factor. The interpretation depends then essentially on knowing the amount of primary extinction, and this can be determined; while secondary extinction can be considered as a second-order effect so long as the reflection takes place from the same face of the crystal and the intensity of reflection remains approximately constant. The same holds good of the true absorption coefficient so long as only a narrow range of wave-lengths, for which the coefficient of absorption changes little, is explored. Its change becomes very important in the transition from one to the other side of the edge, and must then be determined with great care.

Zinc presents certain anomalies because it belongs to the group of anisotropic crystals* for which both the f value and the temperature term in the intensity expression are a function of the orientation of the reflecting lattice planes with regard to the principal crystallographic axis, so that particular difficulties arise in measurements in which the intensities of various reflections are compared, since the interpretation then depends on the knowledge of the various parameters. This point does not interfere with measurements in which the same reflection is used throughout, and the actual knowledge of the numerical value of f is only required for normalizing purposes. Zinc was chosen in the first place because its K absorption edge at 1·2806 A. is conveniently situated and because large single zinc crystals were available. We are indebted to Dr W. Berg for supplying us with a number of specimens. We examined the (0002) reflection; the white radiation was emitted from an X-ray tube with a copper anticathode and a coated filament, run at a constant voltage of 29,000 V. The use of such a filament is of practical importance, because with a tungsten filament a deposit of tungsten on the anticathode gives rise to faint tungsten lines in the neighbourhood of the zinc absorption edge to which we have referred earlier. The copper lines were useful for the process of separating the irregularly scattered from the regularly reflected radiation and for fixing the scale of wave-lengths.

---

* An account of recent measurements of the anisotropic behaviour of zinc can be found in G. E. M. Jauncy and W. A. Bruce(13). See also G. W. Brindley(14).

The method of the slanting slit was used. After interpreting the photographic records by means of a blackening curve and making the allowances discussed in the earlier part of this paper, we obtained an intensity-distribution curve for white radiation scattered from the zinc crystal. This could be compared with the intensity curve covering the same range of wave-lengths and taken with a rock-salt crystal, which in this particular range presents no anomalies since both constituents, sodium and chlorine, have their absorption edges far removed from the absorption edge of zinc. Alternate exposures with rock-salt and with zinc were interposed, so that any irregularities in the running of the x-ray tube could be eliminated. The intensity-distribution for zinc could thus be expressed in terms of the intensity-distribution observed for sodium chloride; any deviation of the zinc curve in the neighbourhood of the zinc absorption edge could then be attributed to the change in the scattering factor of zinc and to the change of absorption for wave-lengths near to the edge.

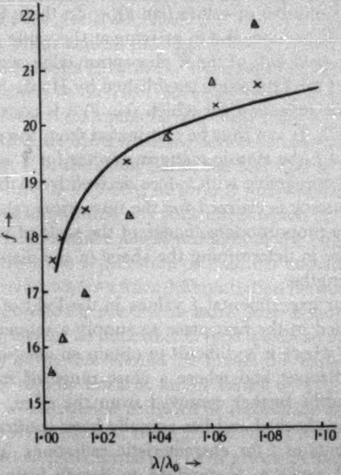


Figure 5. Values of f on the assumptions that the crystal is perfect (△) and that it is imperfect (×).

A detailed account of these evaluations, which involve new measurements of the absorption coefficient of zinc for wave-lengths near to the K absorption edge, will be given in a discussion of Berry's measurements. We limit ourselves here to indicating the main results on the long-wave side. Different interpretations arise according to whether the crystal can be considered as perfect or imperfect in the accepted sense of these terms. If it is perfect, the observed variation of intensity of reflection should be accounted for by a change in f proportional to this variation; if it is imperfect, the scattered intensity should be taken as proportional to f^2, and the change in f should be derived on this assumption. In figure 5 we have entered relative values of f, for a number of wave-lengths, derived from our measurements on each of these two assumptions. For convenience of comparison the values are normalized so that in each case $f = 20 \cdot 0$ when $\lambda/\lambda_0 = 1 \cdot 05$, λ being the reflected wave-length and λ_0 the wave-length of the absorption edge, equal to 1·2806 A. This value of f is that for the particular reflection used, for which $\sin\theta = 0 \cdot 203$, and is derived from tabulations prepared by James and Brindley⁽¹¹⁾ in which values of f are calculated according to Hartree's method for a number of values $(\sin\theta)/\lambda$. In these tabulations no account is taken of anomalous dispersion, and in arriving at the value 20·0 corresponding to $\lambda/\lambda_0 = 1 \cdot 05$ on the long-wave side of the K absorption edge, we have made allowance for dispersion by use of the expression established by Hönl. Since all our measurements refer to the same reflection for which $(\sin\theta)/\lambda$ is constant, the temperature factor is the same for all. It can thus be eliminated from the explicit calculation by normalizing in terms of f, the atomic scattering factor for $T = 0$. This is convenient since this f is directly comparable with values derived from theoretical expressions. No high degree of accuracy is claimed for the numerical value 20·0; any error involves, however, only a proportionate change of the scale of ordinates and is therefore of little significance in determining the shape of the dispersion curve within a small range of wave-lengths.

We may discuss our experimental f values in the light of the assumption that the new method is called in the first place to supply a sequence of values in close proximity of the edge, where it is difficult to obtain an adequate number of values with characteristic radiations and where a close range of measurements is most needed. For wave-lengths further removed from the edge, the wave-mechanical dispersion curve can be checked and has actually been confirmed from a sufficient number of measurements of f for characteristic radiations. We can thus use this part of the theoretical curve as a standard to decide from it whether the zinc crystal approximates to the perfect or to the imperfect type. In figure 5 the wave-mechanical dispersion curve for zinc is indicated by the line. We have calculated it according to Hönl's method;* again we have normalized it so as to make f equal to 20 when $\lambda/\lambda_0 = 1 \cdot 05$. Our calculation takes account of only the principal term in Hönl's expression and not of any dispersion effect of the zinc L shell, the effect of which is only small for wave-lengths in the neighbourhood of the K edge. The curve as traced by us is established in a similar way to the dispersion curve for iron calculated by Hönl and given in his paper. If now we compare the experimental values obtained when $\lambda/\lambda_0 > 1 \cdot 04$ with this curve, we find that the experimental values derived on the assumption of an imperfect crystal agree much better with the calculated curve and that the face of the zinc crystal used in our experiments, which was obtained by cleaving a single crystal at the temperature of liquid air and

[*] It is the term ξ^0 in expression (15) of Hönl's paper⁽⁷⁾ in the *Annalen der Physik*, p. 650.

polishing it, was essentially constituted by a mosaic of crystallites of the imperfect type. The points marked × which are obtained on this assumption thus give a substantially correct interpretation of the observations. If we consider now the range nearest to the edge, it will be seen that the points approximate to Hönl's curve. The experimental points indicate a somewhat steeper rise of f than would be indicated by the curve. The differences which do not exceed 0·2 electron are, however, on the borderline of the experimental error. Only the point nearest to the edge fits Hönl's curve less well. For the immediate neighbourhood of the edge some departure from the behaviour indicated by Hönl's expression must be expected; on the other hand, the point is so near to the edge that its location is less certain; we have inserted this point mainly to show that for wave-lengths in the immediate neighbourhood of the edge f still decreases and that any large departure from the Hönl curve which may occur near to the edge is confined to a region very close to it.

The interest of these results is twofold. On one hand, the checking of the wave-mechanical dispersion curve as such has more significance in a range very near to the absorption edge than for wave-lengths farther away from it, since it is in the neighbourhood of the edge that the differentiation between expressions based on different assumptions is greatest. On the other hand, the fact that in the neighbourhood of the edge, for values of λ/λ_0 up to 1·003, no serious interference is superposed on the smooth dispersive curve shows that any such effect which on theoretical grounds is actually expected in the immediate neighbourhood of the edge is confined to a very narrow region and justifies the expectation of a similar undisturbed approach to the edge on the short-wave side, where a finer differentiation between theoretical assumptions is possible. In regard to the present measurements it should be pointed out that on one hand the assumption of an imperfect crystal is only an approximation; if some correction for primary extinction should be made, the experimental sequence of points would rise more rapidly towards larger values of λ/λ_0. On the other hand, the Hönl curve which we calculated is drawn in accordance with the principal term of his expression, so that the effect of the higher terms is not taken account of. It is not to be expected that the higher terms will be of great significance on the long-wave side and for this reason we have ignored them in our calculation. It would thus appear that comparatively little advantage would be gained by aiming at a higher accuracy on the long-wave side of the edge. It is on the short-wave side that the higher terms, in so far as they modify the distribution law of the virtual oscillators, have the greatest effect on the shape of the dispersive scattering curve. In Williams's treatment this is illustrated by reference to the three curves drawn by him for different assumptions on the distribution law of the virtual oscillators; on the short-wave side these curves show distinct differences in behaviour, whereas they almost coincide on the long-wave side. The same result is brought out from a closer discussion of Hönl's expression. This particular point regarding the distribution law of the virtual oscillators can thus be investigated much better by means of measurements taken on the short-wave side of the edge. There, however, as we have indicated above, both the evaluation of the measurements and the theoretical interpretation depend on possessing very exact values for the absorption coefficient.

It does not seem justifiable to adopt the simplified procedure, used in the present discussion, for the purpose of a quantitative comparison of experimental and theoretical values covering a more extended range of wave-lengths, since in such a case more exact determinations of absorption and extinction are required. The extension of our measurements on the long-wave side up to the value 1·3 of λ/λ_0 indicates, however, a smooth rise of the scattering factor and does not show the sudden inflection observed by Jesse in his measurements on nickel.

REFERENCES

(1) Baxter, A. and Brentano, J. C. M. *Phil. Mag.* (7), **24**, 473 (1937).
(2) Brentano, J. C. M. *Proc. Phys. Soc.* **49**, 61 (1937).
(3) Bragg, W. L., Darwin, C. G. and James, R. W. *Phil. Mag.* **1**, 897 (1926).
(4) Brentano, J. C. M. and Baxter, A. *Z. Phys.* **89**, 720 (1934).
(5) Allison, S. K. and Jesse, W. P. *Phys. Rev.* **49**, 483 (1936).
(6) Jesse, W. P. *Phys. Rev.* **52**, 443 (1937).
(7) Hönl, H. *Z. Phys.* **84**, 1 (1933); *Ann. Phys., Lpz.*, (5), **18**, 625 (1933).
(8) Williams, E. J. *Proc. Roy. Soc.* A., **143**, 358 (1934).
(9) Kallmann and Mark, H. *Ann. Phys., Lpz.*, **72**, 585 (1927).
(10) Brentano, J. C. M. *Phil. Mag.* **4**, 620 (1928); *Proc. Phys. Soc.* **47**, 932 (1935).
(11) James, R. W. and Brindley, G. W. *Z. Kristallogr.* **78**, 470 (1931).
(12) Bruce, W. A. *Phys. Rev.* **53**, 802 (1938).
(13) Jauncy, G. E. M. and Bruce, W. A. *Phys. Rev.* **51**, 1067 (1937).
(14) Brindley, G. W. *Phil. Mag.* **21**, 790 (1936).

THE SENSITIVITY OF PHOTOGRAPHIC FILMS TO X RADIATION AT VERY LOW TEMPERATURES

By JAMES REEKIE, Royal Society Mond Laboratory, Cambridge

Communicated by Dr J. D. Cockcroft, F.R.S., 24 January 1939. *Read* 12 *May* 1939

ABSTRACT. Experiments have been carried out to determine the sensitivity of photographic films to X radiation when exposures are made with the films maintained at temperatures obtainable with liquid nitrogen, hydrogen, and helium. The sensitivity at any temperature T is defined as the ratio of the exposure times required to produce images of equal density (0·25) by exposures made at room temperature and at the temperature T, the intensity of the radiation being the same for each exposure.

It is found that the sensitivity appears to decrease uniformly down to about 100° K., but below that temperature the decrease becomes less rapid, until, at temperatures below 20° K., the sensitivity remains practically constant at a value considerably higher than that found for visible light by Berg and Mendelssohn. The results obtained for X rays are compared with those found by Berg and Mendelssohn for visible light, and it is suggested that the relatively high value found for the limiting sensitivity in the case of X rays indicates that a considerable number of free electrons must be formed by the absorption of the X radiation while the film is at low temperature.

---

§ 1. INTRODUCTION

In the majority of investigations in which use is made of a photographic film as a recorder of X radiation it is generally possible to maintain the film at room-temperature. When experimental conditions are such that this is not the case, a knowledge of the temperature-dependence of the sensitivity of the film becomes of importance. The work of Eggert and Luft$^{(1)}$, which covered a temperature range from −60° C. to +75° C., showed that for most types of photographic film the sensitivity to X rays was approximately a linear function of the temperature, being nearly proportional to the absolute temperature over that range. Jauncey and Richardson$^{(2)}$ however did not find this linear relation, but suggested that the sensitivity might become nearly constant at lower temperatures. The present paper gives the results of an investigation of the relative sensitivity of photographic films to X radiation when the films are cooled to temperatures obtainable with liquid nitrogen, hydrogen, and helium.

§ 2. EXPERIMENTS AND RESULTS

Ilford Ilfex double-coated X-ray films have been exposed, in a suitable camera, to the direct radiation from a Philips X-ray tube with a copper anode, the radiation being passed through a nickel filter. The tube was operated on a.c. at a peak voltage of 35 kv. and passed a current of 15 ma. This secondary current was maintained constant during exposures, by hand regulation of a suitable rheostat in the filament circuit, the operating voltage on the tube being controlled by maintaining the input voltage to the high-tension transformer at a constant value. Fluctuations in secondary current and voltage did not exceed about 2 per cent. Exposures were made either *in vacuo* with the film at room-temperature, or with the film in contact with liquid nitrogen, hydrogen, or helium. The temperature of the film was taken to be the same as that of the liquefied gas, the vapour pressure of which was measured and its temperature consequently known.

In the method adopted to determine the sensitivity it was necessary to know the shape of the curve showing the photographic density of the image as a function of the exposure time to a radiation of constant intensity while the film was maintained at room-temperature. This was obtained by allowing the radiation to fall successively, for different periods of time, on small areas of the film about 0·6 mm. square, while the remainder was shielded from scattered radiation by a lead sheet, the film being shifted slightly between exposures. In this way small blackened portions of the film separated by clear strips were obtained on development, and from the resulting microphotometer trace the photographic densities could be calculated. Processing of the films was carried out by developing in a normal commercial (Kodak) developer for $1\frac{1}{2}$ min. at 15° c. The films were then washed for 1 min. and fixed in the usual way. Blackening due to chemical fog was in all cases practically negligible. Since the different exposures covered altogether only a relatively small area of film (about 6 mm. by 0·6 mm.), they necessarily received almost identical treatment during processing, and the calculated densities could therefore be compared directly.

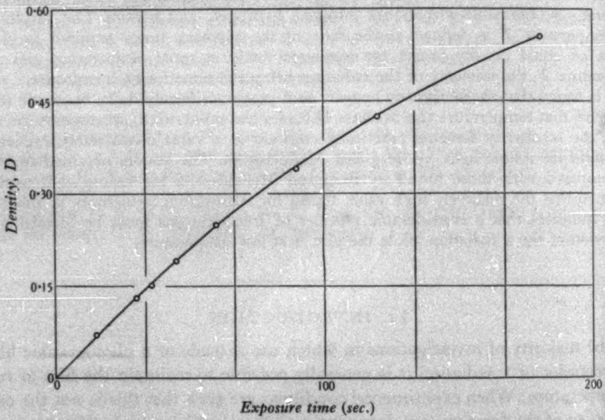


Figure 1.

The result is shown in figure 1 for a film-temperature of 293° K., each point of the curve being the average of measurements made on several films. Density above fog, D, is defined by the equation $D = \log_{10} i_0/i$, where i_0 and i are the intensities of the light transmitted through the clear and through the exposed portions of the film respectively. It is evident that the relation between density and exposure time is, for these films and x radiation, not even approximately linear. A linear relation is usually assumed to hold, and has in fact been obtained in the experiments of Brindley and Spiers(3) and of Jauncey and Richardson(2), for densities less than 1, although in both cases the curve became concave to the exposure-time axis for greater densities. The present results are in general agreement with those given by Bouwers(4), and can be represented by the equation suggested by him, viz., $D = C \log (at + 1)$, C and a being constants and t the exposure time. Also, in agreement with other observers, there is no inertial period evident in the case of x rays. Jauncey and Richardson found that the shape of the curve relating density and exposure time varied with the temperature of the film; this point has not been investigated here since it is not relevant to the determination of the sensitivity by the method adopted.

To obtain the sensitivity the experimental procedure has been to expose the film, at room-temperature, to radiation of constant intensity for two different periods of time, so as to obtain two points on the curve of figure 1. A single exposure to the same radiation was then made with the film maintained at the lower temperature, the time of this exposure being adjusted so as to give a photographic density as nearly as possible equal to that of the more dense image obtained by the exposures at room-temperature. Small differences in the densities of these two images could be taken into account by comparison with the curve of figure 1, which the two exposures made at room-temperature defined with sufficient accuracy for this purpose. The sensitivity, S, of the film at the temperature T is then defined by the ratio of the two exposure times required to produce equal densities (of approximately 0·25), the sensitivity at room-temperature being, of course, unity. Corrections have been applied for the absorption of the radiation in passing through the vapour above the film; these corrections are small, amounting only to a few per cent of the observed density, except in the case of the observations at 80° K., for which somewhat larger corrections are necessary. All measurements have been made on films taken from the same batch.

Figure 2 shows the sensitivity as a function of temperature. No observations were made between room-temperature and 80° K., but it is evident that the results are consistent with an approximately linear variation of sensitivity, proportional to the absolute temperature, over part of this range, as is suggested by the results of other workers(1). Below 80° K. the sensitivity decreases less rapidly, and appears to become practically constant at very low temperatures. The complete curve given in figure 2 refers to the results obtained with the Ilford film. Some observations made with Agfa Sino x-ray film indicate a similar temperature dependence of the sensitivity, but, as shown in figure 2, the relative values of S are somewhat lower than those obtained for the Ilford film. For instance, at a temperature of

4·2° K. the sensitivity of the Ilford film is 0·22, while the value observed for the Agfa film is approximately 0·16. Again, at 80° K. the Agfa film has a sensitivity of about 0·24 compared with 0·32 obtained for the Ilford film. The observed sensitivity appears to be a function only of the temperature of the film at the time of exposure, and is not affected by cooling the film either before or after the exposure is made.

The curve showing density of the image as a function of exposure time in the case of the Agfa film is also essentially similar to that which is given in figure 1 for the Ilford film. In particular, no linear portion is evident, even for small densities.

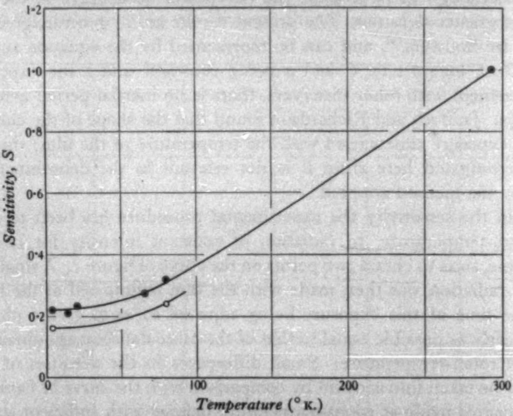


Figure 2. Values observed for: ⊙, Ilford film; ○, Agfa film.

Although the films are brought into contact with the liquid hydrogen and helium and assume correspondingly low temperatures, their mechanical properties on again being warmed to room-temperature appear to be unaffected. It is to be expected, of course, that considerable physical changes, possibly including structural change, would occur in the gelatine at the temperatures used in these experiments. The amount by which any such change may alter the low-temperature sensitivity relatively to that at room temperature is not known, but it is probable that the observations made over the lower temperature range would all be affected to nearly the same extent, and that the values found for the sensitivities at these lower temperatures would not be changed appreciably relatively to each other. In this case the sensitivity would still approach a temperature-independent value at the lowest temperatures.

§ 3. DISCUSSION OF RESULTS

The recent experiments of Berg and Mendelssohn(5) have shown that photographic films still possess an appreciable sensitivity to visible light at temperatures down to 20° K. There are no published measurements on the sensitivity to visible light at still lower temperatures but, according to the calculations of Mott(6) on the production of free electrons by the absorption of light quanta in AgBr crystals, we should expect a marked drop in the sensitivity to visible light when the temperature of the film is lowered to that of liquid helium. It is evident from the present results that no such drop occurs in the case of x radiation, at least down to a temperature of 3·7° K.

In figure 3 Berg and Mendelssohn's results for visible light are compared with those obtained for x radiation. (It should be noted that the definition of sensitivity used by these authors differs slightly from that adopted in the present paper, but this difference will not appreciably affect the curves of figure 3.) Over the temperature range common to both sets of observations the sensitivity falls off in a similar manner in each case as the temperature is lowered. The significant difference between the curves is the considerably higher limiting sensitivity found in the case of x radiation at the lowest temperatures. It would seem that this relatively high sensitivity must be attributed to the formation of an appreciable number of free electrons, liberated by the absorption of the x radiation while the film is still at the low temperature. In the case of visible light, Berg and Mendelssohn's results would indicate that the number of electrons raised directly into the conduction band by the absorption of light quanta is much smaller.

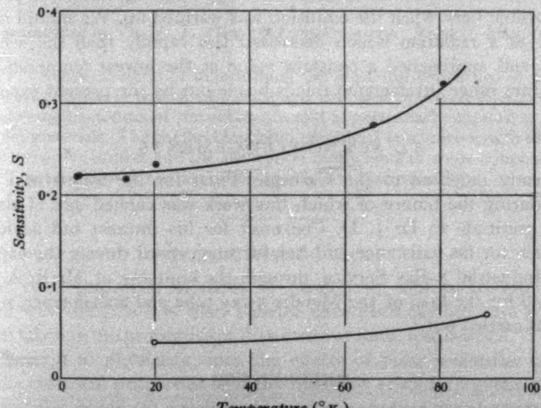


Figure 3. ⊙, sensitivity to x radiation; ○, Berg and Mendelssohn's results for visible light.

According to the model proposed by Gurney and Mott(7) for the formation of the latent image by visible light, the absorption of a quantum of radiation raises an electron of a bromine ion into an excited state. At normal temperatures the thermal energy of the crystal lattice is then sufficient to lift the electron into the conduction band, whence it can travel to one of the sensitivity specks of Ag_2S and so commence the formation of the latent image. For the case of visible light Berg and Mendelssohn⁽⁵⁾ have shown that at low temperatures the above process probably occurs in two stages, the electron freezing in at some point of the lattice while the film is at low temperature and the ionic movement taking place after the emulsion has been warmed up.

In the case of x radiation each absorbed quantum probably distributes its energy among several electrons, either by the formation of Auger electrons or by direct ionization. Some of these electrons will be raised directly into the conduction levels (whence they are free to form negatively charged sensitivity specks) and others into various excited states in the bromine ion. The former electrons would give rise to a sensitivity independent of temperature, while the latter would freeze in while the film was at low temperature and then undergo the same process as in the case of visible light when the emulsion was warmed up. We should thus obtain a sensitivity to x radiation which decreased less rapidly than the sensitivity to visible light and approached a constant value at the lowest temperatures. Over the temperature range investigated this is borne out by the present experiments.

§ 4. ACKNOWLEDGEMENTS

I am greatly indebted to the Carnegie Trust for the award of a Research Fellowship during the tenure of which this work was carried out. I also wish to express my gratitude to Dr J. D. Cockcroft for his interest and advice, and to Dr J. F. Allen for his assistance and helpful suggestions during the experiments. To Philips Industrial x-Ray Service, through the kindness of Mr R. A. Stephen, I am indebted for the loan of the Metalix x-ray tube and transformer, with which the work was carried out.

REFERENCES

(1) Eggert, J. and Luft, F. *Veröff. ZentLab. Anilin. photogr. Abt.* **2**, 9 (1931).
(2) Jauncey, G. E. M. and Richardson, H. W. *J. Opt. Soc. Amer.* **24**, 125 (1934).
(3) Brindley, G. W. and Spiers, F. W. *Phil. Mag.* **16**, 686 (1933).
(4) Bouwers, A. *Z. Phys.* **14**, 374 (1923).
(5) Berg, W. F. and Mendelssohn, K. *Proc. Roy. Soc.* A, **168**, 168 (1938).
(6) Mott, N. F. *Proc. Roy. Soc.* A, **167**, 384 (1938); *Proc. Phys. Soc.* **50**, 196 (1938).
(7) Gurney, R. W. and Mott, N. F. *Proc. Roy. Soc.* A, **164**, 151 (1938).

A NEW METHOD OF DETERMINING THE PROPERTIES OF DIELECTRICS AT CENTIMETRE WAVE-LENGTHS

By M. VELASCO, D.Sc., University of Zaragoza, Spain

AND

G. L. HUTCHINSON, Ph.D., King's College, London,
Keddey Fletcher Warr Student of the University of London

Communicated by Dr H. T. Flint 10 *February* 1939. *Read* 26 *May* 1939

ABSTRACT. A new method of determining the electrical properties of dielectrics at centimetre wave-lengths is described. Although this is a free-wave method it requires only relatively small amounts of the substance and is particularly applicable to measurements on solid materials. The method has been employed at a wave-length of 12·0 cm. to measure the refractive indices and the absorption coefficients of some common laboratory substances. The results obtained are given and briefly discussed.

---

§ 1. INTRODUCTION

In recent years improved methods of generating continuous electromagnetic waves of wave-lengths less than 1 metre have been devised, and great interest has been taken in the properties of dielectrics at these frequencies. The region is of special interest to physicists since the effects of polar molecules considerably influence the electrical properties of many materials at such frequencies.

The quantities which are usually measured in a study of the electrical properties of a dielectric are the refractive index n (or alternatively the dielectric constant ϵ) and the absorption coefficient k. The amplitude of a plane wave travelling in such a medium in the direction x is then proportional to:

$$\exp\left[-2\pi\frac{kx}{\lambda}\right]\exp\left[j2\pi p\left(t-\frac{n}{c}x\right)\right],$$

p being the frequency of the wave, and λ the wave-length inside the medium.

There are two methods by which n and k can be measured at centimetre wave-lengths, the wire-wave method and the free-wave method. The former generally employs a Lecher wire system of which part is immersed in the substance under investigation. The standing-wave system produced is then observed both in the medium and in the air, and from the amplitudes and positions of the voltage antinodes along the wires it is possible to derive values of n and k. The application of the theory can only be approximate, however, whilst inaccuracies on the practical side may be large. The design of the apparatus needs careful consideration, since any irregularities along the wires, the influence of supports, or the proximity of neighbouring bodies is liable to cause spurious effects. The coupling to the oscillator and the effect of the detector are further factors which introduce inaccuracies and fix the lower limit of working with this method at wave-lengths not less than 6·0 cm.

The free-wave method, however, which is really an optical method, is probably of greater use at shorter wave-lengths, but up to the present the various ways in which this method has been employed have suffered from a number of inconveniences. Probably the greatest disadvantage is that only one of the quantities required has been determined—either *n* or *k*, but not both. Thus Cleeton and Williams⁽¹⁾ measured the absorption coefficient of ammonia gas by determining the change in the amplitude of waves passing through a large bag containing ammonia gas. Goldsmith⁽²⁾ has measured *n* only for water by observing the variations of amplitude of a wave reflected from a large tank of water on changing the thickness of the liquid. The theory he employs is only very approximate, as has been shown by Knerr⁽³⁾. The latter measured *n* and *k* for water with a wire-wave method down to a wave-length of 6·5 cm., and at lower wave-lengths he used a free-wave method which gave the product *nk* only.

These free-wave methods have all suffered from the further inconvenience that large quantities of the substance were required, so that their application to a certain extent is limited. The present method, also a free-wave method, suffers from none of these inconveniences. Furthermore, it is most easily applied to solid dielectrics whose behaviour has not hitherto been investigated in the region in question. The method has been developed at a wave-length of 12·0 cm., and values of both *n* and *k* have been determined for a number of common substances by the use of pieces only 5 × 10 cm. in surface area and thicknesses from 0·1 to 3·0 cm.

§ 2. THEORY OF THE PRESENT METHOD

The method is effectively a measurement of the real and the imaginary parts of the dielectric constant, denoted by ϵ' and ϵ'' respectively. The real part gives rise to the induction current in the medium, whilst the imaginary part is related to the conduction current and hence the absorption. These quantities are connected with the generalized refractive index by the relation

$$n^2(1-jk)^2=\epsilon'-j\epsilon'', \qquad \text{......(1)}$$

so that

$$2n^2=\{(\epsilon')^2+(\epsilon'')^2\}^{\frac{1}{2}}+\epsilon' \qquad \text{......(2)}$$

and

$$k\epsilon''=\{(\epsilon')^2+(\epsilon'')^2\}^{\frac{1}{2}}-\epsilon'. \qquad \text{......(3)}$$

It is thus seen that determinations of ϵ' and ϵ'' are sufficient to enable values of *n* and *k* to be obtained.

In this method the material is placed in the path of a parallel beam of radiation having an amplitude E_0, say, incident on the surface of the material. This gives rise in the substance to a certain induced polarization, not necessarily in phase with E_0, of I_0 per unit volume. If we introduce the angle θ by which I_0 lags behind E_0, it is easily seen that:

$$\epsilon' = 1 + 4\pi \frac{I_0}{E_0} \cos\theta, \qquad \text{......(4)}$$

$$\epsilon'' = 4\pi \frac{I_0}{E_0} \sin\theta. \qquad \text{......(5)}$$

Now this induced polarization, being vibratory in character, gives rise to a secondary wave. That part of the wave which would normally be termed the reflected wave, if the size of the material were large in comparison with the wave-length, can be observed by virtue of its interference with the incident beam along the incident direction. The amplitude E_r produced at a distance r away from a slab of the material of area S and thickness t will be given by

$$E_r = \frac{4\pi^2}{\lambda_0^2} \frac{S.t}{r} I_0, \qquad \text{......(6)}$$

where it is assumed for the moment that t is sufficiently small for the change in amplitude and phase of the incident wave passing through the material to be neglected: and that the distance r is sufficiently great for the phase differences which exist between waves arriving from different parts of the surface of the material to be neglected. Thus on measuring E_r for a particular incident wave of amplitude E_0 we can derive I_0/E_0. Furthermore, the value of θ can be measured, as will be shown later, from the positions of the interference maxima and minima produced by the incident and secondary waves. Hence by the use of equations (4) and (5), ϵ' and ϵ'' can be evaluated.

§ 3. EXPERIMENTAL ARRANGEMENT AND PROCEDURE

The first figure shows the experimental arrangement of the apparatus. M indicates the material under investigation and XO denotes the direction of the incident radiation. The transmitter consists of a split-anode magnetron valve working at a wave-length of 12·0 cm., with the aerial—a vertical dipole—set at the focus of a large parabolic mirror. A roughly parallel beam of radiation, polarized in the vertical direction, is thus obtained. The detector D consists of a small tuned aerial system connected to the heater of a vacuojunction, the thermojunction of which is in circuit with a sensitive galvanometer. The deflection of this galvanometer is calibrated in terms of the current in the heater circuit.

In free-wave measurements a certain amount of re-radiation from the receiver occurs, and allowance must be made for this. Subsidiary experiments were performed to determine the re-radiated amplitude at different distances from the receiver in terms of the incident amplitude at the latter. The effect was found to vary inversely as the distance from the receiver, and at 20 cm. the measured ratio of the re-radiated to the incident amplitude was 5 per cent.

This determination is based on the argument that if there were no phase-change the maxima would lie at distances equal to multiples of half the wave-length from the substance. The effect of a phase-change is simply to displace these maxima a distance given by

$$\lambda\theta/4\pi.$$

In general we can say that if a maximum is located at a distance r_M from the substance, then

$$r_M = N\frac{\lambda}{2} - \frac{\lambda}{4\pi}\theta,$$

where N is an integer.

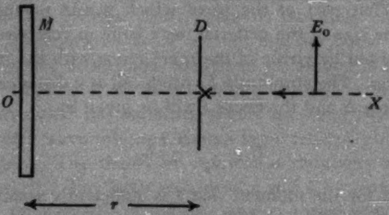


Figure 1 *a*.

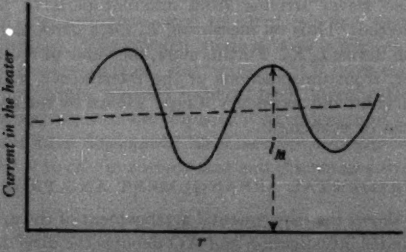


Figure 1 *b*.

When the detector is moved along OX the standing wave system observed is as shown in figure 1 *b*. It is interesting to note that with the limitations mentioned in connexion with equation (6) this curve suffices to give all the information required for a determination of n and k. The positions of the maxima with respect to the material enable one to determine the phase angle θ. Furthermore, the amplitude of the heater current at a maximum i_M is such that:

$$\frac{i_M - i_0'}{i_0} = \frac{E_r}{E_0}, \qquad \text{......(7)}$$

where i_0 and i_0' are the heater currents produced by the incident beam alone, i_0 at the surface of the material and i_0' at the position where the maximum is obtained. They are measured of course without the material in position. Actually i_0 and i_0' are very nearly equal, and would indeed be the same if the beam of radiation were absolutely parallel.

In practice the theoretical limitations already mentioned are overcome by making observations of rE_r/SE_0 and θ for various thicknesses and then extrapolating for an infinitely thin piece of the material. Figure 2*a* illustrates the type of curve obtained, showing the values of rE_r/SE_0 for various values of t. The particular curve shown was obtained for glass. The bending over of the curve with increasing values of t is due partly to absorption and partly to the change in phase of the wave

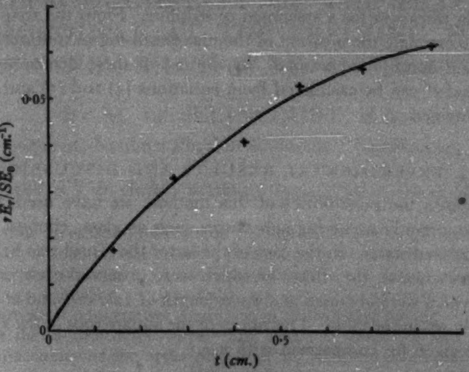


Figure 2*a*.

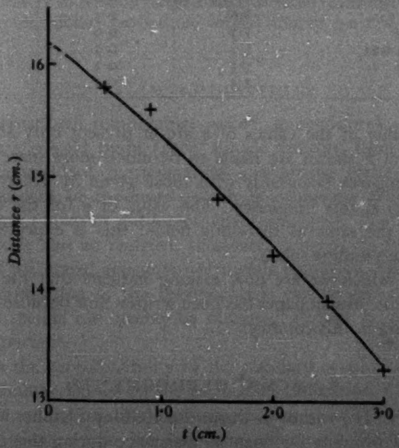


Figure 2*b*.

as it passes through the material. The initial gradient of this curve gives the value of rE_r/StE_0, and hence I_0/E_0 is calculated from equation (6).

Figure 2*b* shows how the position of an observed interference maximum alters with increasing thickness for a specimen of sulphur. From the shape of the curve it is possible to estimate the position of the maximum for an infinitely thin piece of the material and hence θ is obtained. By the use of these determined values of θ and I_0/E_0, ϵ' and ϵ'' can be calculated from equations (4) and (5) and thus values of n and k are obtained.

§ 4. EXPERIMENTAL RESULTS AND DISCUSSION

In investigating the possibilities of this method, we have used, for reasons of convenience, commonly occurring substances such as glass, ebonite, paraffin wax, sulphur and distilled water. In the case of the latter the liquid was held inside thin-walled ebonite cachettes, the effects of which were observed experimentally. The results, all of which were obtained at a wave-length of 12·0 cm. and at a temperature of 20° C., are shown in the table. The values of the square root of the static dielectric constant ϵ_s are given for comparison with n.

Substance	Value of n	Value of k	$\sqrt{\epsilon_s}$
Water	8·6	0·3	9·0
Sulphur	1·3	0·9	2·0
Paraffin wax	1·2	0·3	1·18
Ebonite	1·5	0·2	1·52
Glass	2·5	0·3	2·50

The dependability of the values of n are at present only about 15 per cent, whilst the values of k shown are liable to be much more inaccurate. Our value of n for water compares favourably with those given by Knerr⁽³⁾ (8·85) and by Goldsmith⁽²⁾ (8·95) at this frequency. The value of n for the solids are seen to be approximately the same as for static fields; this is expected in the case of crystalline or semicrystalline solids.

Improvements which we are now making indicate that the accuracy can be greatly increased; the present paper has been written now because one of us (M. V.) is no longer working in this country.

§ 5. ACKNOWLEDGEMENTS

One of us (G. L. H.) wishes to thank the Keddey Fletcher Warr Studentship Board of London University for financial assistance during this research.

REFERENCES

(1) Cleeton and Williams. *Phys. Rev.* **45**, 234 (1934).
(2) Goldsmith. *Phys. Rev.* **51**, 245 (1937).
(3) Knerr. *Phys. Rev.* **52**, 1054 (1937).

THE REPRODUCIBILITY OF THE PLATINUM THERMOCOUPLE AT THE FREEZING POINTS OF GOLD, SILVER AND ANTIMONY

By M. de SELINCOURT, B.A.
Physics Department, National Physical Laboratory, Teddington, Middlesex

Communicated by J. H. Awbery 2 January 1939. *Read in title* 12 *May* 1939

ABSTRACT. The investigation deals with the reproducibility of the standard platinum couple at the three points, viz. the freezing point of gold, silver and antimony, at which it is calibrated for the realization of the international temperature scale from 660° to 1063° c. The factors limiting reproducibility are considered under three heads, namely, electrical measurement, temperature conditions in the furnaces containing the ingots, and homogeneity of the thermocouple wires. The electrical measuring unit is found to be adequate for an accuracy of the order of ±0·01° c., whilst in connexion with the other two factors improvements have been made which render such an increased accuracy just attainable as an upper limit. In conclusion it is considered that the quality of the thermocouple wires, particularly in relation to their state of strain, constitutes the main obstacle to an increase in reproducibility beyond the normal limit of ±0·1° c.

§ 1. INTRODUCTION

On the international temperature scale*, temperatures in the range from 660° c. to the gold point (1063° c.) are deduced from the electromotive force of a standard thermocouple—comprising platinum with platinum containing 10 per cent of rhodium—which is calibrated at three fixed points, namely the freezing points of antimony, silver and gold. The accuracy to which this part of the scale can be realized, therefore, depends on two factors: (*a*) the freezing points of the actual specimens of metals employed, and (*b*) the precision with which the calibration of the thermocouple can be reproduced. Very pure specimens of the metals can readily be obtained, so that in practice the second factor is more important.

As bearing on the reproducibility of the standard couple, attention should be drawn to the recommendations on the method of calibration, which are contained in the specification of the international temperature scale. The more important of these recommendations are to the following effect: (1) The specimen of metal should be brought to a uniform temperature a few degrees above its melting point and allowed to cool slowly with the couple immersed in it through a hole in the centre of the crucible cover: (2) the couple should be mounted in a porcelain tube with porcelain insulators separating the two wires: (3) the depth of immersion of the couple should be such that during the period of freezing the couple can be lowered or raised at least 1 cm. from its normal position without altering the indicated e.m.f. by as much as 1 μv. (0·1° c.): (4) during freezing the e.m.f. should remain constant to 1 μv. (0·1° c.) for a period of at least 5 min.: (5) as alternative to (3) both freezing and melting points may be observed and if these do not differ by more than 2 μv. (0·2° c.) the observed freezing point may be considered satisfactory. Incidentally it may be remarked that conditions (3) to (5) appear to contemplate a reproducibility of the order of ±0·1° c.

---

* Adopted by general agreement in 1927 (1), (2).

A recent inter-comparison⁽³⁾ of temperature scales between the national standardizing laboratories of Germany, Great Britain, and the United States of America, was carried out by circulating a group of six standard thermocouples to be calibrated at the points named. In the result an agreement was found, to 0·1° c. at each point, between the mean calibrations of the group as given by the several institutions, with considerably wider divergences between the values for individual couples.

The purpose of the present investigation has been to study the limits of reproducibility of platinum thermocouples under conditions not differing greatly from those recommended for the international temperature scale, and if possible to improve on the limit previously attained in determinations at the Laboratory, which was considered to be ±0·1° c.

§ 2. SOURCES OF ERROR AND METHODS OF MEETING THEM

The principal sources of error which limit the reproducibility of platinum thermocouples at the three fundamental points fall under the following heads.

(A) *Electrical measurement.* Error may arise in any part of the measuring system, such as the potentiometer, galvanometer, standard cell, reversing switches, etc. or it may result from electrical leakage in the furnace, cold junctions, and the leads from the thermocouple to the instruments, or from parasitic thermal e.m.fs. at any junction of dissimilar metals in the circuit. (B) *Temperature distribution.* Error may arise by departure from the theoretically correct distribution in the surroundings of the hot or cold junctions of the couple. (C) *Homogeneity of thermocouple wires.* A lack of homogeneity may result in an e.m.f. which is not a constant function of the difference in temperature between the hot and cold media. These sources of error and the methods of meeting them are discussed in detail below.

(A) *Electrical measurement.* The precision aimed at in the electrical measurements was ±0·05 μv., or about 1 part in 100,000 of the e.m.f. of a standard couple at the gold point, a figure more than sufficient to cover a possible tenfold increase in the overall accuracy of the calibrations previously obtained at the Laboratory. In addition to the selection of a suitable instrument, this involves the elimination of all electrical leaks and parasitic e.m.fs. which might reach this order, and the provision of a standard cell for checking the potentiometer current which can be relied on to at least ±5 μv.

Measuring instruments. The measuring instrument selected was a potentiometer of the Diesselhorst type, supplied by Messrs Otto Wolff of Berlin and used in conjunction with a Moll galvanometer. This potentiometer is so designed as to be nearly free from parasitic thermal e.m.f., and is supplied with an oil-immersed double-reversing key, which simultaneously reverses the potentiometer current and the potential circuit. The current through the potentiometer is regulated by balancing a standard Weston cell across a special resistance, supplied in a separate box, so that to obtain the value of any e.m.f. in terms of the Weston cell, the ratio of the resistance of this box to that of the principal coils in the potentiometer is of fundamental importance.

As a result of several calibrations, this arrangement was found to be satisfactory, with the following qualifications. (1) The temperature coefficients of the resistances of the potentiometer and current regulating box were found to have different values, namely 0·000018 and 0·000012 per 1·0° C. respectively; thus even a simultaneous change in the temperatures of the two instruments would involve a change in the fundamental ratio of 6 parts in 1,000,000 per ° C., while a difference in temperature between the two would cause a change of more than 1 in 100,000 per ° C. The two instruments being air-immersed and mounted in separate boxes, they could only be used to the accuracy required if the temperature of the room remained steady to at least ±0·25° C. The room in which they were used was controlled thermostatically, but it was found impossible to stir the air sufficiently to maintain these conditions. Accordingly the two instruments were remounted as one unit, and immersed in a bath of stirred oil; in this way their temperatures could be kept equal and steady to the necessary degree. (2) The galvanometer key fitted on the instrument was found to develop sufficient heat in constant use to give rise to a thermal e.m.f. exceeding the required limit of accuracy of 0·1 μV. To get rid of this uncertainty another key, consisting of a tilting glass tube with internal mercury contacts, immersed in the stirred oil bath, was introduced into the circuit, and this was found to be satisfactory. (3) The double reversing key supplied with the potentiometer was found to have a variable resistance on reversal. This defect was corrected by reinforcing the contacts, which were formed by copper leaves bearing on copper blocks, with additional leaves of phosphor bronze, without introducing them into the circuit, and by substituting clock oil for paraffin oil in the key.

With the above-mentioned alterations the apparatus was satisfactory to well within the limits required, being limited only by the sensitivity of the galvanometer, which at 3 m. from the scale gave a deflection of 2 mm./μV.

Leads. The leads from the thermocouple to the measuring unit are of great importance, since there is no means of eliminating any parasitic e.m.f. which arises in any part of the circuit between the thermocouple junction and the reversing key. In the original design they were therefore in unbroken copper from the key to the cold junctions. Test experiments on cold junction immersion in ice showed, however, that the wire employed, 14 s.w.g., conducted sufficient heat to affect the temperature of the cold junctions; the cold junction connexions were therefore made with a light copper wire, immersed in ice in the usual way, which was in turn connected to the heavy copper leads; to prevent the occurrence of thermal e.m.f.

from differences in the copper, the copper-copper junctions were also ice immersed, or enclosed in special lagged junction boxes, to ensure that they were at the same temperature. The insulation of these leads was found on test to be ample, but there remained a serious source of error in the form of direct electrical leaks from the current supplying the furnaces. Refractories have very poor insulating qualities at high temperatures so that considerable leakage is possible from the mains to any earth available in the thermocouple circuit. To reduce this risk to a minimum, alternating current was used exclusively for heating the furnaces. Otherwise, the only cure was complete isolation from earth of the entire remainder of the circuit. Except in very damp weather, when observations were suspended, the resistance to earth was not distinguishable from infinity.

Standard cell. Of the electrical unit, there remains the standard cell, on whose accuracy the whole measurement of e.m.f. depends. The cells, of the standard Weston type, were supplied by the Electricity Department of the Laboratory, and from the work of that department they are known to be constant to a few parts in 10,000,000. They have, however, a temperature coefficient of e.m.f., which at normal room-temperatures (15° C. to 20° C.) amounts to $-0{\cdot}00003$ or $-0{\cdot}00004$ V. per 1° C., the e.m.f. reaching a maximum at 3° C. To maintain an e.m.f. to the required constancy of ± 5 μV., at room-temperature, the cell must be kept steady to at least 0·1° C.; at 0° C., however, the temperature coefficient is only 0·00001 V. per ° C., so that a temperature constant to 0·3° C. would suffice. In view of this, and of the ease of maintaining a constant temperature of 0° C., it was decided to maintain the cells at this temperature. Accordingly two cells were immersed in oil, kept dry by a desiccating agent and sealed in a glass envelope surrounded by ice. An electrically-operated selector-key served to give connexion to either cell.

(B) *Temperature distribution.* The temperature of fusion of a solid is probably realized best by forming an intimate mixture of the solid and liquid phases, such as is ordinarily used for observing the ice point. This method is impracticable in the case of metals at high temperatures, and the normal procedure, as summarized for example in the introduction to this paper, is to cool slowly from slightly above the melting point, or heat slowly from slightly below the melting point, a cylindrical ingot of metal, on the axis of which is situated the thermocouple or other thermometer. To obtain a satisfactory calibration with this arrangement, it is obviously desirable that the outward or inward conduction of heat by the thermocouple should be negligible and that the only gradient of temperature in the ingot should be radial.

When an ingot is melted under these conditions, the course of events, taking into account only the phenomena of the specific and latent heat of the metal and its conductivity, might be expected to be as follows. The thermocouple should show a steady rise in e.m.f. until the melting of the outermost shell of the ingot checks the inward flow of heat. The e.m.f. should then continue to rise very slowly up to the instant of melting of the last shell of metal in contact with the thermocouple sheath, after which it should rise abruptly. A similar course should be followed on freezing. In each case only the last reading before the abrupt break in e.m.f. would correspond strictly with the change of state, though many readings before this would no doubt approximate closely to this final value.

The course of events imagined above is complicated, during both melting and freezing, by the incalculable effects of convection in the liquid metal, and of undercooling during freezing. These two factors, quite apart from the experimental difficulty of obtaining the desired temperature distribution, would render uncertain the production of melting or freezing curves of stereotyped forms, such as are indicated above. In the circumstances it is hardly possible to do more than apply certain general tests to the observations in order to decide whether they can be regarded as giving a satisfactory calibration. Thus it may be laid down that (1) the reading should remain sensibly constant for a specified time, at least 5 min., or for some considerable fraction of the total time during which melting or freezing is taking place, the time of steady reading not necessarily occurring at any particular stage in either process; (2) the raising or lowering of the junction of the couple by an appreciable amount should produce no effect on the reading, thus proving the absence of an axial gradient due to either a faulty distribution in the ingot or conduction along the thermocouple, and the absence of parasitic e.m.f.*; (3) a good agreement should be found between calibrations carried out under different conditions, for example between those obtained during melting and freezing or between those obtained during either melting or freezing with different rates of heating or cooling as the case may be.

In the present investigation use has been made of all three criteria and more particularly of (1) and (2), and the aim has been to secure that none of the tests should reveal an uncertainty exceeding the highest accuracy desired, namely $\pm 0{\cdot}1\ \mu$V.

Realization of the desired distribution. We now proceed to consider the possibility of realizing in practice the two desiderata already indicated for the obtaining of satisfactory melts and freezes. The first of these—the reduction to a minimum of conduction along the thermocouple—offers little scope for variation, but the other—the suppression of all gradients other than radial gradients in the ingot—opens a big field in furnace design. It is true that the effects are somewhat interlinked, but it is convenient to consider them separately.

Effect of conduction along thermocouple. Some conduction along the thermocouple is inevitable, but it may be reduced by the use of wires, for the couple and leads, as fine as is consistent with adequate strength, and by thin and close-fitting insulators and sheaths. Further, the effect of conduction may be minimized by increasing the length of couple immersed. In the case of the cold junction the difficulties presented are slight. Thus the temperature-difference between the medium and the atmosphere being of the order of 15° C., finely divided ice can be maintained without difficulty in equilibrium with water, and the depth of immersion is limited only by the size of the ice bath and the length of the thermocouple wires. Temperature gradients along the thermocouple wires can thus be kept very small, and the temperature of the junctions readily maintained within 0·001° C. of the freezing point. At the hot junctions, on the other hand, the difficulties are considerable, since the temperatures of the media, i.e. freezing antimony, silver or gold, differ from room-temperatures by 600° C. to 1000° C. Large temperature-gradients must therefore occur between the ingots and the outside air, while the quantity of metal available severely limits the depth of immersion. There is thus a distinct possibility of the conduction along the thermocouple distorting the radial gradient in the ingot. Apart from the factors already mentioned, the materials of which the sheath and insulators are made also affect the conduction of heat. Here the choice is very limited. In the specification of the international temperature scale porcelain is recommended as the material both for sheaths and insulators. In the experiments to be described below silver was melted in a vacuum and antimony in a controlled atmosphere. The use of silica for the sheath was found to be particularly convenient in those cases, and it was adopted also for gold with no apparent disadvantage.

---

* Discussed below under the heading (C) (homogeneity of thermocouple wires).

Temperature-control in furnace. A considerable number of experiments were made before a final design of the furnace and auxiliary gear was reached. Trials were first made with furnaces of conventional type consisting of cylindrical tubes heated by resistor elements, and provided with supplementary heaters for the adjustment of temperature distribution. By careful manipulation of these heaters it was found possible to reduce end loss to a considerable extent, and to improve the reproducibility of calibration to ±0·5 μv. But even for this accuracy, which does not represent a very large advance on that previously attained, the manipulation of these furnaces proved to be troublesome, and the required degree of temperature uniformity could not be reproduced with certainty. In the final design of furnace an improved smoothing effect was sought by interposing a heavy metallic envelope between the heating element and the ingot.

The first experiments on these lines were made by immersing a small ingot of gold in a bath of liquid silver. This bath is contained in a cylindrical fire-clay crucible, 10 in. long and 3½ in. in diameter and covered by a mabor lid. The gold ingot, 3 in. long and 1 in. in diameter, is contained in a closed-end pythagoras tube 15 in. long, which reaches from the outside of the furnace to within 2 cm. of the bottom of the silver bath, so that the surface of the gold ingot is 5 in. below that of the silver bath. The assembly is substantially the same as that shown in figure 1 for use at the melting point of silver. In order that the silver bath should be maintained at a few degrees above or below 1063° C. as required, its temperature was controlled by a thermostat working in conjunction with a platinum resistance thermometer immersed in the bath. Facilities for stirring the bath were also provided, but preliminary experiments showed that if the silver was stirred so as to be at a uniform temperature, a downward temperature-gradient occurred along the axis of the gold ingot, owing to the surplus loss of heat upwards; by discontinuing the stirring, and lowering the crucible assembly in the furnace so as to maintain an upward gradient in the silver, the downward gradient along the axis could be counteracted and it was found possible to reduce the gradient along the axis to 0·1 μv./cm. or less. When this had been done, it was found that the other two conditions for a reliable calibration, namely that during the change of state the e.m.f. should remain constant, and that different cooling curves with the same couple should yield the same result, were satisfactorily fulfilled with the new furnace; but this improvement was realized only after parallel improvements on the thermocouple mounting, as detailed below, had also been effected.

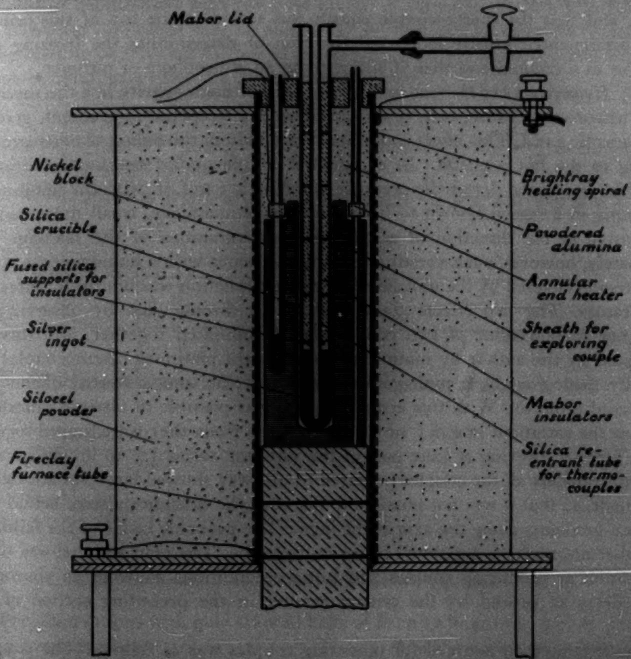


Figure 1. Furnace assembly of silver and antimony points. In the original gold furnace a bath of liquid silver of the same dimensions is used in place of the nickel block, and the crucible containing the ingot is an open ended pythagoras tube, of the same dimensions as the silica crucible shown. Otherwise the furnaces are identical.

Since the stirring of the bath was found to be undesirable, there was no reason why the bath should not be replaced by a block of solid metal and this was accordingly done for the experiments on the melting points of silver and antimony. The metal selected for this purpose was nickel, which, though of a comparatively low conductivity, has the advantage that it only slowly deteriorates by oxidation, a hard crust of oxide being formed which protects the remainder of the metal. The use of a metal which remained solid made it possible to turn off and insulate a shoulder, which accommodated an annular end-heater. This heater, which consisted of platinum wire 0·75 mm. in diameter wound on a steatite ring, was used in place of external windings to compensate for loss of heat up the crucible opening.

The other details of the apparatus used for the silver and antimony points are shown in figure 1. It will be seen that a silica crucible was used for containing the metal, and that the thermocouple sheath was sealed to the top of this crucible. This arrangement greatly facilitated the plan of determining the freezing point of silver in a vacuum and that of antimony in an atmosphere of nitrogen.

(C) *Homogeneity of thermocouple wires.* Lack of homogeneity in a thermocouple wire, whatever its cause, will, if associated with a temperature-gradient, give rise to a parasitic e.m.f. This effect is of importance only in the region of steep gradient leading to the hot junction. In a mounted couple under working conditions it would be feasible to detect the presence of parasitic e.m.f. if, with the hot junction remaining in a region known to be of uniform temperature, small movements of the couple in the direction of a gradient gave variations of the reading. Unfortunately, however, such variations might also occur with a homogeneous couple if the region covered by the movement of the hot junction was not itself uniform in temperature. The two effects could perhaps be disentangled by observing the difference in behaviour of two couples rigidly mounted together, and this was attempted by the author by mounting two standard platinum thermocouples in a specially made four-bore pythagoras insulator of the same external diameter as those used hitherto. With this arrangement it was possible to observe differences between the variations in e.m.f. of the two couples when they were moved together in the furnace, proving at least part of the variation to be due to inhomogeneity; but unfortunately all the couples prepared in this way showed large variations with movement, so that it was not possible to disprove the existence also of actual temperature-gradients along the axis of the ingot. The probable cause of the failure of these thermocouples is suggested below on page 709. The plan adopted was therefore to aim at producing couples and furnace conditions showing an absence of both effects as proved by the criteria set out in the preceding section of this paper.

The first method adopted for mounting couples was as follows. The couples, 1·3 m. in length, were annealed electrically for 30 min. at a temperature of 1500° C., and the hot-junction ends were then threaded through twin-bore porcelain or pythagoras insulators, 18 in. long, 4 mm. in diameter, and with 0·7-mm. bores. The remainder of the wire to the cold junction was insulated with systoflex, the two arms being bound together to prevent the formation of an inductive loop. This method of mounting involved the threading of long lengths of wire through narrow holes in the insulators after the annealing had taken place, a process quite capable of giving rise to small strains in the wires. Such strains would be susceptible to reannealing on prolonged immersion in the furnaces, and might thus give rise to gradual modification of the calibration. To overcome this objection the couples were subjected to a second period of annealing after they had been inserted in their insulators, so that they were relieved from strain as far as possible in their finally assembled state.

§ 3. RESULTS OF CALIBRATIONS AT THE THREE BASIC POINTS

Specimen curves obtained with thermocouples, reannealed as described above and used in conjunction with the final design of furnace, are shown in figure 2. The following comments may be added. In no case was an experiment passed if during melting or freezing the test of moving the junction along the axis showed a change in e.m.f. exceeding 0·1 μv./cm. of movement.

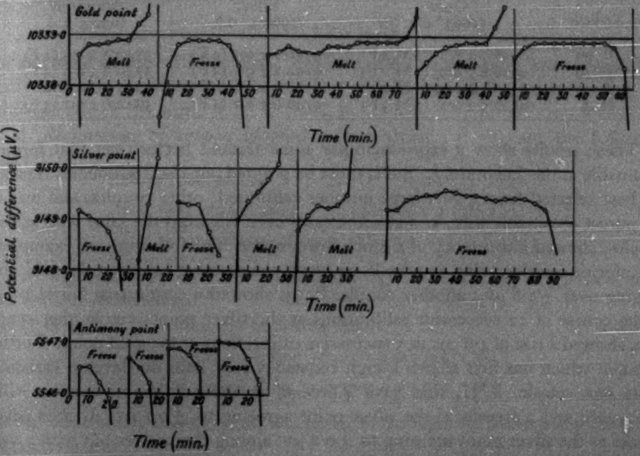


Figure 2. Specimen cooling curves, obtained with couple no. P. 29.

The observations with gold showed halts of from 10 to 40 min. during which the reading remained constant to ± 0·1 μv. The agreement between melts and freezes also fell within this limit.

With silver, the cooling curves do not in general show such good flats, as can be seen from the specimen curves, and in particular the melts are not reliable, showing in most cases a steady rise of about 0·1 μv. per minute. But provided the temperature gradient was not greater than 0·1 μv./cm., and the air pressures over the ingot not greater than 0·001 mm. of mercury, calibrations for the best couples were constant to ± 0·1 μv., which is the same accuracy as obtained for the gold point.

In the case of antimony it was not found possible to obtain a melting curve with a halt slower than about 0·4 μv./min., and the calibrations were in consequence confined to freezing points. Further, owing to the very large undercool exhibited by this metal, which might be as great as 30° c., the flat portion of the cooling curves after the undercool were in general very short. The results show a reproducibility rather less than that obtained for gold and silver, namely ± 0·2 to 0·3 μV.

The calibration figures for the four couples selected as being the best of all those tested are set out in table 1 below. Each set of figures gives the mean value of the e.m.f. in μV.: the extreme range of departure from the mean: and, in brackets, the number of melts and freezes taken into account.

Table 1. Calibration in microvolts of four best couples at the three basic points

Couple	Gold (1063° c.)	Silver (960·5° c.)	Antimony (630·5° c.)
P. 27	10326·7 ± 0·1 (2M 2F)	9140·0 ± 0·1 (4F)	5540·8 ± 0·2 (4F)
P. 29	10338·8 ± 0·0$_5$ (3M 2F)	9149·2 ± 0·1 (4F)	5546·8 ± 0·3 (4F)
P. 30	10338·2 ± 0·0$_5$ (2M 2F)	9148·5 ± 0·1 (4F)	5543·2 ± 0·1 (4F)
H. 3	10327·7 ± 0·1 (1M 3F)	9140·6 ± 0·1 (4F)	5542·6 ± 0·1 (4F)

These results show a reproducibility some tenfold better than that formerly obtainable at the Laboratory. But it must be pointed out that the couples represent the best selected from a very large number calibrated; other couples, also made of the purest obtainable wire, and in some cases even cut from the same wire as these couples, showed a constancy of a much lower order. The following are examples of the inferior couples. One couple, P. 25, gave 7 successive values at the gold point ranging over ± 1·8 μV.: another couple, P. 33, showed a continuous fall of 4·8 μV. in the course of 12 successive calibrations at the silver point: yet another couple, E. 2, showed a rise of 4·6 μV. in 5 successive calibrations at the gold point. Further, a couple which has first shown a high constancy may later on develop variations. Thus one couple, P. 31, first gave 4 freezes at the gold point coming within ± 0·1$_5$ μV., and 2 freezes at the silver point agreeing to ± 0·1 μV. It then gave 4 freezes at the silver point agreeing to ± 0·2 μV. among themselves but with a mean value as much as 2·0 μV. below the former value. As all these results were obtained with the same furnaces, the varying degree of reproducibility shown by the different couples is to be attributed to inhomogeneity in the element wires.

To investigate this inhomogeneity and its modification on annealing, experiments were carried out on the following: (1) couples such as those referred to above which had shown exceptionally bad agreement; (2) couples which on previous calibrations had given good agreement; and (3) wires artificially strained for the purpose.

(1) *Reannealing of couples of poor reproducibility.* The couple P. 25 referred to above was reannealed and recalibrated. A summary of the results obtained at the gold point, before and after this process, is given below in terms of the notation of table 1.

Before reannealing 10327·5 ± 1·8 (7F)
After reannealing 10335 ± 0·9 (7F)

Similarly treated, the couple P. 33, which had previously shown a progressive decline of 5 μv. in 12 experiments at the silver point, now gave an increase of 10 μv. over the lowest value with a range of only ±0·2 μv. in 3 determinations.

In other experiments, reannealing also raised the mean e.m.f. and improved the reproducibility. But a contrary case can also be quoted. P. 31, already referred to, had given the following results at the silver point, in terms of the notation of table 1, the gradient of e.m.f. complying in both cases with the conditions stated above.

9117·5 ± 0·1 (2F)

9115·5 ± 0·2 (4F)

After reannealing it showed a steady rise in 4 successive readings represented by

9115·5 ± 2·3 (4F)

with a marked deterioration in e.m.f. gradient effect so that the second anneal had greatly reduced the reproducibility of the couple, in addition to increasing the inhomogeneity of the wire.

(2) *Reannealing of couples of good reproducibility*. P. 30 gave the following result:

Before reannealing	10335·3 ± 0·1 (3F 1M)
After first reannealing	10338·5 ± 0·1 (2F 2M)
After second reannealing	10338·25 ± 0·05 (2F 2M)

The high reproducibility was retained throughout, but a rise of 3 μv. occurred after the first reannealing.

The couple H. 3 yielded the following result:

Before reannealing	10323·7 ± 0·3 (3F 1M)
After reannealing	10327·7 ± 0·1 (2F 2M)

The reproducibility, already good, was raised to the highest standard and the e.m.f. increased by 4 μv.

The experiments summarized under (1) and (2) may be understood by assuming that the wires in question were in a condition of strain, uniformly or otherwise, and that this strain was modified on reannealing. A wire exhibiting a uniform strain should have a good reproducibility in addition to showing no variation of e.m.f. when moved along the axis of the ingot, since the additional e.m.f. due to the strained condition of the wire would be dependent only on the temperatures of the hot and cold junctions, and so would form a constant addition to the total e.m.f. of the couple. Reannealing, whether it partially modified the strain or entirely removed it, would thus affect the total e.m.f., but would not alter the reproducibility unless it introduced inhomogeneities into the wire; these could be detected by moving the junction along the axis. This effect is exhibited by P. 30. On the other hand, a wire exhibiting a non-uniform strain will have a poor reproducibility, since the parasitic e.m.fs. arising in the non-uniform parts of the wire will depend on the temperature conditions in their own neighbourhood, which will vary throughout the experiment, as well as between different experiments made under different conditions, and of course if the inhomogeneous parts are moved to different parts of the furnace. On reannealing, the total parasitic e.m.f. will be affected as before, by a general modification of the state of strain over the whole wire, but in addition to this the magnitude of the parasitic e.m.fs. due to the inhomogeneities will be reduced, and reproducibility will be improved. This is the effect observed in couples P. 25, P. 33 and H. 3.

(3*a*) *Artificial straining of couples.* The foregoing observations are based on the assumption that wires of similar material, but in a different condition of strain, will in fact exhibit a thermal e.m.f. in relation to each other. To verify this assumption, and to investigate both the magnitude of the effect and its modification on annealing at various temperatures, the following experiments on the straining of wires were carried out. For these experiments new annealed wires of both pure platinum and platinum-rhodium alloy were divided into equal lengths and rejoined by twisting to form couples which having similar arms should give zero e.m.f. at all temperatures if the two arms were actually in an identical condition. One arm of each couple was artificially strained by bending the wire through 180° round a glass tube of 8 mm. in diameter, a tension of about 500 g. being required to pull the wire round the tube. The open-circuit e.m.f. of the resulting couples was then observed with the hot junction maintained at 1000° C., both before straining and after a number of successive strains, with the following results (the polarity of the strained wire being indicated):

(1) Pure platinum wires

	1 New wire	Old wire, reannealed at 1600° C.	2 New wire	3 New wire
E.m.f. before straining	+ 0·5	+19·0	+ 9·0	+ 2·0
E.m.f. after straining once	+85·0	+21·0	+ 3·5	+11·5
E.m.f. after straining twice	—	+ 2·5	+110·0	+15·0
E.m.f. after straining 3 times	—	+18·0	+ 8·5	—
E.m.f. after straining 4 times	—	—	+ 9·0	—
E.m.f. after straining 5 times	—	—	+ 7·0	—

(2) Alloy wires

	1 New wire	Old wire, reannealed at 1600° C.	Old wire, reannealed at 1600° C.	2 New wire	Old wire, reannealed at 1600° C.
E.m.f. before straining	+ 1·0	− 2·5	+11·0	+ 2·0	−10·5
E.m.f. after straining once	− 9·2	−10·2	+ 5·5	+10·5	−16·5
E.m.f. after straining twice	−10·2	−13·6	+ 1·5	+ 1·5	−21·0
E.m.f. after straining 3 times	—	—	− 1·0	−17·5	—
E.m.f. after straining 4 times	—	—	− 6·5	—	—
E.m.f. after straining 5 times	—	—	+ 4·0	—	—

The e.m.f. observations were repeated after 5 min., and again after 24 hr., with similar results in each case, of which the following are examples:

E.m.f. immediately after straining	10·3	13·5	85·0
E.m.f. after 5 min.	9·0	12·7	83·0
E.m.f. after 24 hr.	9·0	12·7	83·0
E.m.f. after 72 hr.	9·0	12·7	83·0

Thus the effect of strain on a wire is to introduce a thermal e.m.f. with respect to an unstrained wire. In the case of pure platinum, the strained wire appears to be positive; although subsequent strains may reduce the e.m.f., they do not render it negative with respect to the unstrained wire. The magnitude of the e.m.f. when the hot junction is at 1000° C. can in some cases be as high as 100 μV., though it is more usually of the order of 10 or 20 μV. In the case of the alloy wire the strained wire can be either positive or negative with respect to unstrained wire and subsequent strains on the same wire can alter the sign of the e.m.f. The magnitude of the e.m.f. for a hot junction temperature of 1000° C. does not appear to exceed 20 μV. These effects are not temporary, but persist for a number of days at least, although a slight reduction of the e.m.f. usually occurs within a few minutes of straining.

(3*b*) *Annealing of artificially strained couples.* The couples were then annealed at a number of temperatures between 800 and 1600° C., the e.m.f. being observed after each anneal. The usual effect of annealing is to reduce the e.m.f. as shown below:

	Platinum	Platinum	Alloy	Alloy	Alloy	Alloy
E.m.f. unstrained	0·0	2·0	+ 1·0	− 2·7	−10·5	+ 1·5
E.m.f. strained	83·0	15·0	−10·5	−12·5	−21·0	−17·5
Annealed 5 min. at 800° C.	71·0	9·5	− 5·5	−11·0	—	—
Annealed 60 min. at 800°	71·0	9·5	− 2·5	− 6·5	−19·0	—
Further annealed						
60 min. at 1000°	65·0	—	—	− 3·7	−16·3	−10·5
60 min. at 1200°	65·0	—	—	− 2·5	−15·5	−10·5
60 min. at 1400°	30·0	—	—	—	−10·5	−10·5
60 min. at 1600°	16·0	7·5	—	—	−10·5	−10·5

The effect of each anneal is permanent, subject to a slight regression shortly after the annealing has been completed, particularly if the annealing was not sufficiently prolonged for the full effect to take place. The length of time necessary for this is usually upwards of 1 hr. though, in platinum annealed at low temperatures, as little as 5 min. may suffice. Examples of this are given below:

Pure platinum wire	800° C.	1000° C.	1400° C.	1600° C.
Unannealed	83	71	65	35
Annealed 5 min.	71	65	49	30
Annealed 60 min.	71	65	35	16
Annealed 120 min.	71	65	32	15

Alloy wire	800° C.	1000° C.	1400° C.
Unannealed	12·6	17·6	3·8
Annealed 5 min.	11·2	12·0	3·8
Annealed 60 min.	7·2	8·5	2·9
Annealed 120 min.	5·5	8·0	2·9

In a few exceptional cases the effect of annealing was the opposite to that shown above, namely to increase the parasitic e.m.f. If these exceptional cases (which may be due to the annealing of residual strain in the so-called unstrained arm) are ignored, the normal effect of annealing is to reduce the e.m.f. due to strain. The extent to which this reduction of e.m.f. takes place varies with each case, but it will be observed that even for an annealing temperature of 800° C. the effect is considerable, and may exceed 50 per cent. Annealing at progressively higher temperatures above 800° C. reduces the e.m.f. still further until it reaches a residual value which does not appear to be affected by further treatment; for different cases this state may be reached at any temperature from 800° C. to the maximum possible annealing temperature; and the residual e.m.f. may have any value from that exhibited by the unstrained wire to as much as 50 per cent of the maximum.

These effects will affect the practical use of platinum thermocouples in various ways. In the first place, a thermocouple wire if sufficiently severely strained exhibits a thermal e.m.f. with respect to an unstrained wire even after prolonged annealing, so that it is reasonable to assume that, in many cases, one or both of the arms of actual thermocouples will still exhibit a certain amount of residual strain. The largest parasitic e.m.f. is exhibited by the platinum arm, and this is always positive with respect to the unstrained wire. Since platinum is negative with respect to the alloy wire, the result of a strained platinum arm in a platinum-rhodium thermocouple will be to reduce the total e.m.f.; and since the effect of annealing is to reduce the parasitic e.m.f., the annealing of a couple in which the platinum wire is strained will in general raise the total e.m.f. The effect of strain on the alloy wire can be of either sign, but appears more commonly to be negative with respect to unstrained wire. Where it is negative the effect of annealing will, of course, be to increase the e.m.f. of the couple.

These strains will not affect the reproducibility of the couple if they are uniform over the whole length of the wire, but superimposed on them there may be strains of a non-homogeneous nature, which, as has been shown above, can account for the variations in the e.m.f. calibration. The magnitude of the parasitic e.m.f. due to an inhomogeneity in the wire will be of a smaller order than the additional e.m.f. due to a continuously strained wire, since the two ends of the inhomogeneity, being close together, will not be at widely different temperatures. Consequently the effect of annealing may be to modify the total e.m.f. by an amount considerably in excess of the variations.

It is also possible for inhomogeneities to be set up in the wire after it has been put into service, and for these to be modified without reannealing. In the ordinary course of use the element wires of a couple will be subject to slight movement relative to the insulators in which they are contained, owing both to actual handling of the free wires outside the insulators, and to the differential expansion of the wires and fireclay insulators when they are placed in the furnace. Since the space available is somewhat restricted, this movement will give rise on one hand to friction between the thermocouple wires and the walls of the insulator, and on the other hand, if the wires become jammed at any point, to tension or bending according to the direction of the relative movement. Support is lent to this view by the author's experience, referred to previously, with the mounting of two couples in the same insulator. In order to get four bores into one insulator of the standard size, it was necessary to reduce the diameter of the bores to 0·6 mm., and if the foregoing observations are correct this would be expected greatly to increase the difficulty of maintaining the wires in a strain-free condition. It was in fact observed that none of the couples so mounted was comparable in uniformity with the best of those mounted in two-bore insulators, the variation with movement being of the order of 0·5 to 1·0 μv./cm., and that this variation as well as the actual calibration of the couples altered considerably after the couples had been put into service.

From the experiments on annealing it appears that considerable modification of freshly strained wire can result from annealing at temperatures as low as 800° C. A couple in use will be exposed for long periods to temperatures of this order, precisely in that part of their length in which inhomogeneity will give rise to a parasitic e.m.f., so that if inhomogeneity does arise, it will be liable to continual modification, and hence small variations in the calibration of the thermocouple will occur.

§ 4. CONCLUSIONS

It is shown in the experiments summarized above that with proper precautions the standard platinum-rhodium thermocouple is capable of a reproducibility of the order of ± 0·1 μv., equivalent to 0·01° C., i.e. there is a tenfold increase in the accuracy previously obtainable at the Laboratory. With the introduction of certain modifications the electrical measuring apparatus and the temperature conditions in the furnace can be relied upon to this accuracy; the principal difficulty in attaining it lies in the selection of sufficiently homogeneous wire for the thermocouples, only a fraction of the total number of thermocouples prepared by the author having reached this standard. The presence of strain, original or acquired, is seen to be the main cause of uncertainty. This uncertainty might be reduced with a greater knowledge of the effect of annealing, or by the design of special mountings to keep the wires in a strain-free condition, though such mountings might prove so cumbrous as to nullify the peculiar advantages of thermocouple measurement.

§ 5. ACKNOWLEDGEMENTS

The author desires to record his indebtedness to his colleagues Dr F. H. Schofield and Messrs A. Grace and C. R. Barber for valuable advice and assistance.

REFERENCES

(1) *Proc. 7th General Conference of Weights and Measures* 1927, p. 56.
(2) Revised text in *Trav. Bur. Int. Poids Mes.* **19** (1934).
(3) *Collected Researches of National Physical Laboratory*, **24**, 115 (1933).

SOME RECENT DEVELOPMENTS IN BRITISH SURVEYING INSTRUMENTS

By MAJOR-GENERAL M. N. MACLEOD, D.S.O., M.C.
Ordnance Survey, Southampton

The Thomas Young Oration, delivered 24 March 1939

§ 1. INTRODUCTION

THE subject I have chosen for The Thomas Young Oration is a meeting ground of the sciences of optics and of surveying. As it is not possible in the time available to cover the whole ground, I must confine my descriptions to developments which seem to me to be of special interest, by reason either of their novelty, or of their importance to the surveyor; and I must further limit my descriptions to optical as opposed to mechanical features.

Much of what I have to say is concerned with aerial surveying, for this is an important recent development which has led to the construction of surveying instruments of an entirely new kind, some of them veritable robot surveyors and really wonderful products of optical ingenuity and mechanical workmanship. Before embarking on the subject of air-photographic surveying instruments, however, I would like to refer to some recent developments of the ground surveyor's principal instrument, the theodolite, which has not been, and is never likely to be, superseded by the aerial camera.

§ 2. THEODOLITES

Thirty years ago it seemed that finality in the design of theodolites had been reached, and that further improvements would be in matters of minor detail only. However, in 1921* or thereabouts, Herr Heinrich Wild (now of Heerbrugg, Switzerland, but then a member of the Zeiss firm) introduced a new instrument in which the graduated arc was engraved upon glass instead of upon metal, and in which the reading of the arc was effected in a novel manner, so effective and yet so beautifully simple that one wonders why no one had ever employed it before. In previous theodolites, in order to eliminate eccentricity, it had been the universal practice to read the arc separately at two points 180° apart, and to adopt the mean as the final observation. Wild conceived the idea of bringing the two reading points together into a single microscope by means of an optical system. The interpolation between the divisions of the graduated arc was then made with a single micrometer, which could be read without leaving the observing position, and which automatically gave the exact mean of the two readings.

---

* The first patent is dated 1908.

An answer from the country of Ramsden and Sisson to this challenge was eminently desirable; and the gauntlet thus thrown down by Herr Wild was taken up by Brigadier Winterbotham (then Chief of the Geographical Section of the General Staff) who arranged (in 1926) for a meeting to be held at Tavistock at which the principal manufacturers of British surveying instruments, and a number of other persons interested in surveying, were invited to examine and discuss these new instruments, and the question of theodolite design generally. The result of the Tavistock meeting has been the manufacture by Messrs Cooke, Troughton and Simms of the so-called Tavistock theodolite, which uses the new method of reading in another and, I think, even better way.

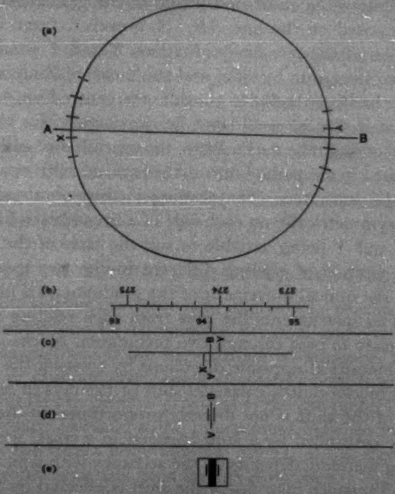


Figure 1.

The new method of reading may be explained as follows. Suppose figure 1*a* to represent a graduated circle. Imagine a line *AB* ruled right across through the centre; the reading of one end of *A* on the arc would be, say X° and a fraction, and of the other end *B*, Y° and a slightly different fraction, where $Y = 180 + X$.

The final reading will be *X*, plus the mean of the two intercepts *AX* and *BY*, i.e. $\frac{1}{2}(AX + BY)$, figure 1 *c*. This mean reading is measured directly with an optical micrometer which moves the images of the graduations *X* and *Y* by equal amounts in opposite directions. In the Wild (and Zeiss) instruments the images of *X* and *Y* are brought together by the optical systems so that they appear as in figure 1 *b* or 1 *c*. If each of two adjacent opposite graduations is displaced optically by equal amounts towards the imaginary line *AB*, figure 1 *c*, until they are exactly in line with one another, each will have travelled exactly one-half the distance *XY* in figure 1 *c*, that is to say a distance equal to $\frac{1}{2}(AX+BY)$. The micrometer scale and the two sections of the arc are brought into the same field of view and both can be read without moving from the observing position.

In the Tavistock instrument the optical system brings two opposite parts of the arc, i.e. the ends of the line AB, together in such a manner that the two ends A and B can be imagined as superimposed upon one another to form a single line represented by the central line in figure 1 *d*. If the graduation and centring of the arc were quite perfect, the optical system would superimpose any two adjacent pairs of graduations 180° apart so that they would appear in the field of view as a second line parallel to the line representing AB. The reading could be effected by an optical micrometer which shifted the combined image of the graduations, e.g. X and Y, until it was superimposed on the line AB. If, however, there were any error in centring or graduation of the arc, the graduations X and Y would appear not as a single line but as two lines side by side, and the mean reading would be obtained when these two lines had been made to straddle the centre line AB as in figure 1 *d*. This arrangement, as it stands, would not be satisfactory, for the X and Y lines might not be clear of, or might be too far from, the central line: a tilting block of glass is, therefore, interposed in the path of one of the rays, and this enables the images of the graduations to be separated. The reading is then obtained by bringing the images of X and Y symmetrically on each side of a broad black line as in figure 1 *e*, the separation of X and Y being variable to suit the taste of the observer.

Not only is the method of reading different in the two instruments, but the method of effecting the optical movement of the graduations is also quite different. In the Zeiss and Wild instruments the rays from each reading point are brought together and then made to pass through two rectangular blocks of glass which can be tilted by equal amounts in opposite directions by rotating the micrometer head. The tilting of these blocks causes a parallel shift of the rays in opposite directions, but as the amount of the shift is not directly proportional to the tilt of the blocks the tilting has to be controlled by a stud engaged in a spiral groove, and the precision of the reading therefore depends on the accuracy with which the spiral groove can be cut. In the Tavistock instrument the shift is effected by two wedge-shaped pieces of glass moved along a slide by an accurately cut rack and pinion, in a manner similar to that used in the well-known Barr and Stroud range-finder. In this case the displacement of the graduations is directly proportional to the movement along the slide, and therefore to the rotation of the micrometer head. This design, in which the accuracy depends on movement along a straight line, seems to me definitely better.

All these theodolites enable an angle to be read by estimation to a fraction of a second. The graduation of the arc borders on the miraculous. In the $3\frac{1}{2}$-in. Tavistock instrument the graduations are 0·0003 in. wide, and about 0·01 in. apart, this spacing representing 20 min. of arc. A single second read on the micrometer is therefore represented by a length of 1/120,000 in. on the graduated circle, yet the results obtained under test show that the errors of division are much less than one second.

Before leaving the subject of theodolites I would like to mention two other useful British developments. One is the magneto-theodolite, and the other is Dr Hunter's shutter eye-piece.

§ 3. THE MAGNETO-THEODOLITE

The magneto-theodolite is an instrument designed for the observation of accurate magnetic bearings. It is a combination of theodolite and magnetometer in which the change-over from angle-measurement of terrestrial objects to measurement of magnetic bearings can be made in a few seconds. The instrument makes use of Sir Frank Smith's principle of fluid suspension in that the magnet (consisting of two strips of cobalt steel) is attached to an annular brass float immersed in a liquid.

The magnet system is placed centrally over the horizontal arc while the telescope is mounted eccentrically, the change-over from magnetic to direct (ordinary) observations being made simply by turning a milled head.

The instrument was constructed for the Ordnance Survey by the Cambridge Instrument Company to the design of Mr H. L. P. Jolly. It is the first of its kind, and it is hoped that it will prove useful for mining as well as for magnetic surveys: indeed for any type of survey in which accurate and quickly obtained magnetic bearings are required.

§ 4. DR HUNTER'S SHUTTER EYE-PIECE

The object of the shutter eye-piece is to improve the precision of time-determination by star-observations by eliminating personal and systematic errors. An illuminated scale against which the star-position is estimated, and a shutter, controlled by a clock or chronometer, which alternately occults and exposes the star under observation at known times is provided in the eye-piece. The position of the star at each appearance has only to be interpolated between the lines of the eye-piece scale, and by the time a large number, say 24, of star flashes have been thus positioned the personal error has been largely eliminated.

Tests of the shutter eye-piece indicate that with its aid it is possible to make observations for longitude equal in accuracy to those for latitude, i.e. equivalent approximately to an error in absolute position of about 30 to 40 ft.

§ 5. AIR PHOTOGRAPHY

With these remarks as an introduction I will pass on to the subject of air surveys and air-survey instruments.

As everyone knows, a map shows the features of the earth's surface much as they would appear to an observer looking down from above. If the ground were flat, if there were no distortion in the photograph due to the lens or shutter, and if the camera could be held exactly vertical, an air photograph would appear to obtain, in an instant, information which might take a surveyor on the ground many days, and perhaps weeks, to collect.

Unfortunately the above theoretical conditions can very seldom be realized exactly. It is consequently exceptional to be able to use an air photograph for systematic mapping by mere enlargement or reduction. The photograph must be regarded only as a picture with certain definite geometrical properties from which the surveyor can extract the information he requires. Moreover it is a picture at a much reduced scale, and though this may be an advantage, in making the features to be mapped more easily accessible, it postulates a greater refinement of measurement than would be required at the full scale on the ground itself.

§ 6. CAMERAS

The first consideration in aerial photographic surveying is a suitable camera; and since it may take an hour or more of flying to carry the camera to the point at which work commences, the camera should allow the greatest possible area to be photographed before the aircraft has to return to its base. Practically, then, the camera should make use of film and not glass plates.

For the same reason the angular field of the lens should be as wide as possible, but since the use of the photograph assumes that it is a true perspective of the ground the field must be limited to the width that can be covered by the lens without perceptible distortion.

The exact determination of the geometrical properties of this perspective require that the optical axis should coincide with the perpendicular to the plate from the rear node of the lens, and that the principal distance (the length of this perpendicular) should be accurately known. In British air-survey cameras this requirement is met by pressing the film just before exposure against a flat glass plate rigidly fixed in the focal plane of the lens. On this plate are engraved fiduciary marks, or in some cases a fine réseau, which are photographed upon the film at each exposure, and which indicate the position of the principal point. Air-survey cameras should be calibrated by bringing the principal point, as defined by the pressure plate, into coincidence with the optical axis, and by determining the principal distance. The lens, camera body, and the pressure plate should, in effect, form a single unit which should remain unaltered in dimensions after calibration.

The shutter must be capable of giving an exposure short enough to avoid any blurring or distortion of the image, but focal-plane shutters which draw a slit in a blind across the plate and thereby expose each part of the plate in succession are unsuitable for precise air-survey work. British air-survey cameras are generally fitted with a Louvre shutter behind the lens.

The standard British service air camera is known as the F. 24. It is a film camera taking a 5 × 5-in. picture capable of being fitted with lenses of 6 in. and 4 in. focal length.

Other British air cameras are the Eagle cameras made by the Williamson Manufacturing Company. These also are film cameras and they are made in several sizes. The Eagle IV takes a picture 18 × 24 cm. The Eagle VI takes a 5 × 5 in.

picture and is specially designed to provide an optical system which, after calibration, can be operated, serviced, and mechanically adjusted, without disturbance of the lens or the pressure plate.

§7. LENSES

It is only during the last six or seven years that lenses specifically designed for air survey have been produced in this country. The Ross wide-angle "Xpres" lens of 7-in. focal length is the first of these deserving of special mention. This lens gives sensibly sharp definition at an aperture of F/4 over a square plate of 7-in. side. The maximum distortion measured radially from the centre is at 35·6°, and is then no more than 0·03 mm. At 35° it is zero.*

Excellent though this lens is, the light travelling along the oblique rays to the margins of the plate is very much less than that falling upon the centre. Subsequently, therefore, the firm of Ross produced another lens in which the performance in this respect was improved without any appreciable sacrifice of its other attributes. This lens is known as the "Xpres E.M.I." (extra marginal illumination) lens.

The latest production of the Ross firm is a wide-angle survey lens designed to work at an aperture F/5·5 and having an angular field of more than 90°.

The field of these new lenses must, I think, be the practical limit which can be reached with a single lens. Any further increase in covering power can only be obtained by the use of multi-lens cameras. A 9-lens camera, or rather a 9-camera assembly, was used by Scheimpflug, the pioneer of air photography, in his attempts to photograph from balloons in 1910, and about ten years ago a Munich firm produced a 9-lens camera covering an angular field of 135°, and quite recently another 9-lens camera assembly has been produced in the United States.

§8. THE 7-LENS AIR CAMERA

In 1933 the Air Survey Committee decided to try and evolve a camera with a slightly less angular field (120°), but with fewer lenses. The original design was got out by Lieutenant J. S. A. Salt, R.E., with the assistance of Mr Booth and Mr Stringer of the Royal Aircraft Establishment, and the camera was constructed by Messrs Barr and Stroud.

It consists of a cluster of seven lenses, one at the centre and six at the corners of a regular hexagon. The focal length is a little under 2 in., and the rays entering the six outer lenses are deflected by prisms, fixed immediately in the front of the lenses, which have the effect of bending the rays entering the side lenses through an angle of 45°. In front of the prisms is a yellow light-filter which also forms a circular glass cover for the whole assembly. The lens cone is designed to fit the film-holder and magazine of the standard F. 24 camera. The glass pressure plate is replaced by a metal mask which divides up the exposed portion of the film in such a manner that the seven separate pictures project, after rectification, into an exact square. During exposure the film is pressed against the mask by a pad which is automatically released when the exposure is complete and the next section of film is moved into position.

---

* See the 1935 Report of the Air Survey Committee.

In the rectifiers used for foreign cameras the printing of the oblique pictures is done by rotating each picture in succession into the appropriate position in the rectifier. With the 7-lens camera the negative is put into a special rectifier in which all seven pictures are printed simultaneously, and the six oblique pictures are rectified during printing to the scale of the central picture, which is magnified about 1¼ times. The design of this rectifier is exceedingly skilful and compact. Simultaneous printing of all the photographs involves a certain complication not only in the number and arrangement of the projection lenses but also in the lighting arrangements, since the amount of light required for the side pictures is many times that required for the projection of the central picture, but it is bound to save a great deal of time.

The enormous saving in photography and flying which can be effected by the use of wide-angle and multi-lens cameras can be seen from the following table:

Flying height 15,000 ft.	F. 24 camera Xpres lens	7-lens camera
Side of each picture	2·9 miles	9·8 miles
Area of each picture	8·4 sq. miles	96 sq. miles
Number of photographs per thousand square miles	410	32
Miles of flying per thousand square miles	460	140

§ 9. PLOTTING APPARATUS

Having now dealt briefly with the apparatus for taking an air photograph, I will pass on to the instruments, etc., required for converting it into a map.

An air photograph shows the ground in two dimensions only, but the third dimension (i.e. height) can be measured by combining, either stereoscopically or otherwise, two views taken from different standpoints. Figure 2 shows a hill M photographed from two air stations A_1 and A_2. The position and height of M could be determined by placing positives of the photographs in two projecting cameras, of the same dimensions as the aerial camera, fixed to the end of a bar A_1A_2, the length of which is decided by the scale of plotting. In the aerial camera, however, the lens is set for focus at infinity, and in order to obtain a sharp image from these projectors on a screen representing the ground at the same scale as the bar, a lens of different focal length would have to be used, or the pictures would have to be projected piecemeal, an auxiliary focusing device being interposed in the path of the optical rays. With the screen at M_0 a point such as M, lying above the ground plane, will have two images, at M_1 and M_0 respectively, but if the screen is moved towards the projectors these two images will converge until they coincide when the screen is at M. The projection on the screen in this position is obviously the map position of the point M, while MM_0 represents its height above the datum level. A French plotter, the Gallus-Ferber, is based directly on these principles. Automatically focusing auxiliary-lens systems, controlled by guide rods, throw a

Rectified and unrectified prints from the 7-lens camera.

By permission of the Air Ministry

sharp image of a small portion of each picture on to a movable screen. The two pictures are not projected simultaneously but are thrown on to the screen alternately. The effect is that when the two images of any point are not coincident the picture appears to shake from side to side. The screen can be raised or lowered until the shaking ceases, when the position of the point can be plotted.

The majority of automatic plotters, however, depend on stereoscopic methods of combining the two pictures. The principles of stereoscopic plotting may be explained as follows. If the two pictures are viewed in a suitable stereoscope, instead of being projected on a screen, a single picture will be seen in relief, giving a clear impression that M is nearer to the observer than M_0. Suppose now that a strip of transparent material is placed over each photograph and a mark, say a cross, is made on each strip exactly over S_1 and m_1 respectively. The two marks will fuse and be seen as a single mark coinciding with the point M in the stereoscopic view. If the mark centred on m_1 is then moved slowly from m_1 to m_0, the two marks will not go out of fusion, but the fused mark will appear to move away from the observer, until it coincides with M_0 in the stereoscopic view when the actual mark is centred over m_0 on the photograph. In other words the lateral movement of the mark m_1 to m_0 gives a measure of the height MM_0. Convergence of the two marks makes the fused or floating mark appear to rise; divergence makes it appear to fall or recede. Much the same effect is produced if the marks are kept stationary and the separation of the photographs is altered. Small movements of one of the photographs relative to the other in the direction of the base A_1A_2 do not affect the fusion of either the marks or the photographs, but only cause an apparent movement of the fused impression of the former.

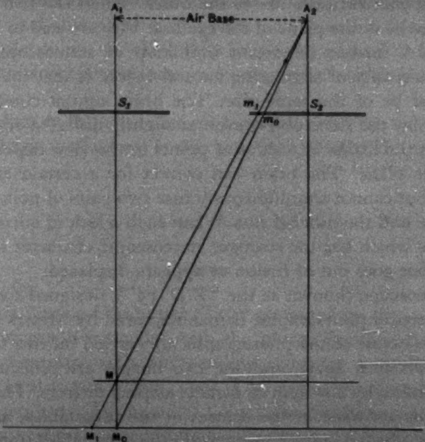


Figure 2.

§ 10. STEREOSCOPY

The stereoscopic impression of depth is purely relative, and is limited to the field of distinct vision, i.e. about 1° around the point of fixation. It is not possible to get a satisfactory stereoscopic impression of an object of such a height that the angle subtended at either eye much exceeds 1°; hence the greater the range of relief the more nearly parallel must the visual rays be.

Easy stereoscopic fusion is only obtained when each pair of images lies in the same plane as the two eyes. The brain can keep in fusion two points which are only slightly out of the eye plane, but could not, without inclination of the head, fuse two pictures one vertically above the other. When the two components of a stereoscopic image lie in the plane of the eye base they are said to be in stereoscopic correspondence. A further important limitation of stereoscopic reconstruction, imposed by the necessity of simulating natural vision, is that the separate pictures of an object must be of the same size. The brain cannot comfortably fuse two pictures which show the same object even at slightly different scales.

Easy fusion of the marks as well as of points in the view depends on both lying strictly in the eye plane. The brain can correct for a certain amount of lack of correspondence, but cannot simultaneously fuse two pairs of points of which one is in correspondence and the other is not. When such a lack of correspondence exists that pair of points which has the stronger stereoscopic character remains in fusion, and the other either goes out of fusion or appears displaced.

A simple stereoscope (known as the "Z.D. 15"), designed for the stereoscopic examination of vertical photographs, is manufactured by Messrs Barr and Stroud. The prints (or positives) of the photographs are carried on two turn-tables placed at a convenient distance apart, and the two images are reflected into the eyes, without magnification, by a system of parallel sloping mirrors. The principal points of the photographs are fixed at the centres of the turn-tables, and the setting in correspondence is effected simply by rotating the turn-tables until the line joining the two principal points on each photograph of the pair is parallel to the eye base. Multiple marks are provided in the form of a grid engraved on a glass plate placed over each photograph and of which the lines run at 45° to the eye (or air) base. When the photographs are in correspondence the paired lines of this grid fuse to give the impression of a horizontal network, the height of which can be varied by altering the separation of the grids, this movement being given by a screw with a graduated micrometer head at the right-hand side of the instrument.

If one or other of the photographs is taken with a slight tilt, or is placed a little out of line in the stereoscope, the correspondence setting will be affected and, although the photographs may still be fused easily enough, the grid will no longer appear as a horizontal network. When the grid is brought approximately to the ground level, the fused crosses will split so that one diagonal appears nearer to the observer than the other. The reason for this splitting, which can be used to check correspondence setting, can be seen from figure 3.

Suppose that two photographs have not been set quite correctly in correspondence and that, owing perhaps to a slight rotation of the right-hand photograph, the line joining *P*, *P′* (the two images of a point) is not parallel to the base, it can be seen that the lines *YY′* will be at a lesser separation, and *XX′* at a greater separation, than the fused image of the point. The *Y* lines will therefore appear nearer, and the *X* lines farther, from the observer. If the grid is moved in the direction of the arrows, until the *X* line appears to be in coincidence with the ground level, that is to say when *P* lies exactly on *X* and *P′* on *X′*, the *Y* line will be seen fused but apparently appreciably above the ground.

It will be observed that stereoscopic measurement implies the simultaneous alignment of two marks on two separate points. Its advantage lies largely in the fact that two operations are performed simultaneously, automatically, and yet very accurately. This dual alignment is of course only possible when the movements of the marks and the points on which they have to be aligned lie in the same plane; the condition of stereoscopic reconstruction requires that this plane should contain the eye base also.

Methods of plotting from vertical photographs (with the aid of the "Z.D. 15" stereoscope) were first worked out in this country by Major Hotine, R.E., under the auspices of the Air Survey Committee. They are based on the assumption that the centre of the photograph coincides with the plumb point and consequently that rays drawn from the centre to the surrounding points on the photograph will define the correct horizontal angles at that point. Two such rays plotted from the ends of the base will consequently fix any other point. The advantage of a technique using vertical or nearly vertical photographs is that the aerial navigation is simplified, that the plotting apparatus is simple and relatively inexpensive, and above all that the method is easy to teach and lends itself well to mass production.

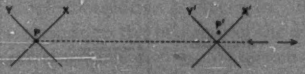
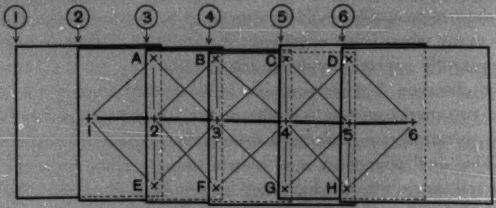


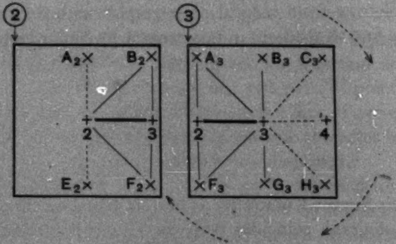
Figure 3.

With the aid of gyroscopic aircraft stabilizers it is now quite practicable to take air photographs in straight strips with the correct overlaps (which should be between 50 and 60 per cent) in the fore-and-aft direction, and with no deviations from the vertical exceeding half a degree. A skilled aircraft crew can ensure that the whole of any given area shall be covered in this manner with a minimum of wasted flying time; and since flying time is expensive, this is an important consideration.

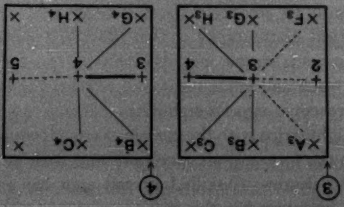
With photographs taken in this manner each photograph will show near its right and left margins the centres (called "principal points") of its neighbours in the strip on each side. The first stage in the plotting consists in fixing the relative positions of the principal points of a strip, and four selected points in each overlap, by a sort of triangulation on the scale of one selected photographic pair; see figure 4 *a*. A skeleton plot of each strip having been prepared in this manner, the next step is to bring all the plots to a uniform scale, and to determine this scale, and the orientation of the whole, by reference to a few independently fixed control points suitably distributed over the area. This fixes the position of the centre of each photograph and the length and orientation of each air base. The third stage is the plotting of the detail by intersecting rays from the two principal points in each



(*a*) Diagram of air triangulation.



(*b*) Setting for the overlap 2–3.



(*c*) Setting for the overlap 3–4. The turn-table carrying photograph no. 3 has been rotated 180°.

Figure 4.

overlap. The last stage is the plotting of the contours, by means either of stereoscopic parallax measurements or of form lines drawn in with the aid of the stereoscope and heighted from aneroid or other measures of height made on the ground.

With good photography, small random tilts are probably not the chief sources of error in the control plots; errors in identifying the points, stretching or shrinking of the film after exposure, and limitations of draughtsmanship, are the chief sources of inaccuracy. In order to eliminate errors of these three classes a special comparator, designed by Captain E. H. Thompson, R.E., has recently been constructed for the Ordnance Survey by the Cambridge Instrument Company.

§ 11. THE THOMPSON COMPARATOR

The instrument is a combination of a measuring stereoscope and an accurate co-ordinate measuring machine. The photographs are placed on two rotating turn-tables each mounted separately on a carriage, called the "*X* carriage", movable in a direction parallel to the stereoscopic eye base. The stereoscope is mounted on another carriage, called the "*Y* carriage", moving in a direction exactly perpendicular to the former. The two *X* carriages can be locked together, but each can be given an independent movement, so that a differential movement of the turn-tables can be made in the *X* direction. The mechanical features of the instrument will not be described in detail, though they are in some respects the most interesting and the most characteristic points in the design.

It has been explained above that if the photographs are taken with suitable overlaps each will show, near its right and left margins, the principal points of the adjoining photographs on the strip on each side, and a narrow strip of the ground will appear on each of three consecutive photographs. By intersecting selected photograph points on this strip from the two ends of the air base common to each pair of photographs, an aerial triangulation can be carried out on the photographs (see figure 4) and the relative positions and lengths of the several bases can be computed. If, however, the basal angles of this triangulation were measured directly, errors would be introduced by irregular distortion of the film base, or of prints. The angles are not, therefore, measured directly, but, in accordance with the practice of precise astronomical photography, are calculated from rectangular co-ordinates measured with the aid of an accurate non-distorting réseau inscribed on the pressure plate of the air camera, and recorded photographically on each negative at exposure. Errors arising from film distortion are therefore limited to the expansion or contraction within a single square of the réseau.

As the use of a réseau is a characteristic feature of this instrument, it is worth noting, in passing, that most continental air-survey cameras do not allow the use of a réseau as they do not employ a glass pressure plate. Other features are the method of ensuring exact identification of the same points on all three photographs, and the method of changing over from one stereoscopic pair to the next. These are best explained by describing the method of use.

A pair of photographs are placed on the turn-tables and centred; the tables are then rotated to bring the base lines on both photographs into coincidence with the line of the *X* movement, and their separation is varied until the floating mark in the stereoscope, centred over each principal point in turn, appears as though resting on the ground. The setting can be made with greater precision in this way than by attempting to identify a particular point of detail under the mark, since it depends not only on the centring of the mark but on the stereoscopic impression received from all the surrounding detail.

The floating mark is formed by the junction of two dissimilar T-shaped marks, one in each eye-piece; see figure 5. When the two photos are exactly in correspondence and each half of the mark is centred correctly over its respective photo point, both horizontal and vertical lines will appear collinear, and the mark will appear as a cross resting on the ground. Any lack of correspondence in the photographs, or error in identifying the same point on both photographs, will immediately show itself by the mark dividing, and one half appearing out of line with or separated from the other. This type of mark gives therefore a very sensitive setting.

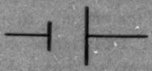


Figure 5.

The base lines having been set in correspondence, the next step is to select the intersected points. It is sufficient to place the floating mark in a suitable position on one photograph and identify stereoscopically the point indicated by the mark on the other by bringing the floating mark down on to the ground in the view. Where slight tilts exist there may be some lack of correspondence at the edges of the photographs, even when the base setting is correct. This will be revealed by the splitting of the floating mark and must be eliminated without alteration of the setting. This is done by means of parallel plate micrometers which enable the image of either of the photographs to be shifted in the *Y* direction. After each setting has been completed, the distances of each point, from the nearest lines of the réseau, the co-ordinates of a suitably placed corner of the réseau, and the micrometer reading, are read off on the scales provided for the purpose, and booked. From these measures, and the known dimensions of the réseau, the basal angles can be obtained and the triangles solved. Some short cuts in the computation can be taken, and most of it can be done by machine.

The next step is to record the position of the selected points so that when one photograph of the pair is removed and the next inserted the same points will be correctly identified on the latter. This is done by means of specially designed optical markers. In order to introduce the next photograph in the strip with the minimum of trouble, the stereoscope is made symmetrical to right and left about the central axis, so that after removing one photograph it is only necessary to rotate the photo left in the stereoscope through 180° and place the next one on the vacant turn-table. This is equivalent to measuring up the next stereoscopic pair upside down; see figures 4*b* and 4*c*. The optical markers, figure 6, which can be clamped to the turn-table consist of a fine cross engraved on a glass plate 1 which is carried in a short movable tube 2. (This tube can be rotated so as to set the mark at an angle with the cross formed by the stereoscope graticule.) Approximately central in the tube and fixed in position is a lens 3 and at the end of the tube is a 45° prism 4. The effect of the optical system of the marker is to form a real image of the cross at *m*, and a virtual image at *m*′. The tube can be adjusted so that this virtual image lies in the plane of the paper.

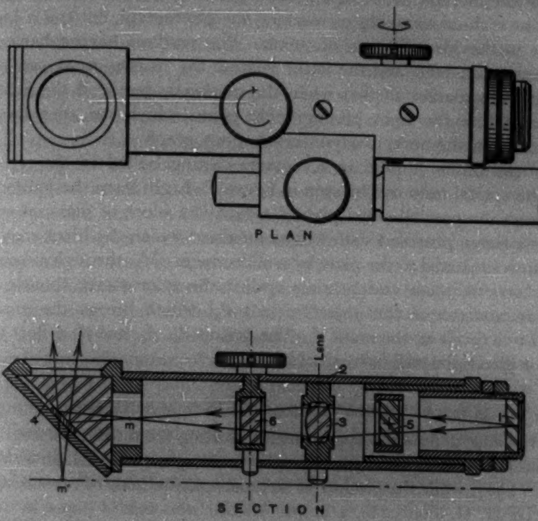


Figure 6.

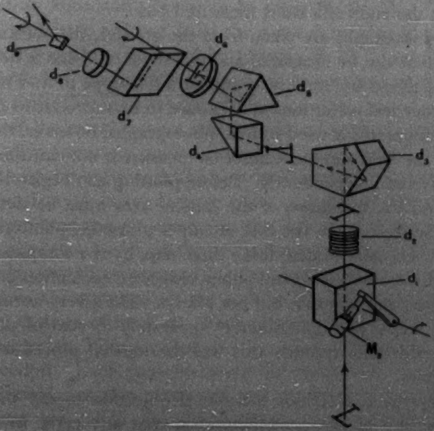


Figure 7

This optical method of bringing the mark into exact coincidence with the selected point without touching or marking the photograph ensures a high degree of accuracy in the identification of points. The markers having been placed in position on the turn-table and the latter rotated, the *floating* mark can be centred in turn over each marker so that when the marker is removed the floating mark remains directed at the right photographic point, which can then be identified stereoscopically on the next, newly inserted, photograph.

The optical unit is housed in a separate casting bolted to the top of the *Y* carriage. The optical train can be seen in figure 7. Light from the photograph first passes through the parallel-sided plate d_1 which is a block of glass, 50 mm. thick, mounted in a frame provided with trunnions about which the block may be tilted. The rotation is imparted to the plate by a micrometer, M_2, through a weighted arm attached to one trunnion and bearing against the micrometer spindle. Directly above the micrometer is the objective lens d_2, which brings the image of the photograph to a focus on the mark d_6. The prisms d_3, d_4, and d_5 deflect the rays in the required directions and form an erect image.

§ 12. AUTOMATIC PLOTTING MACHINES

It now remains to describe apparatus for plotting from tilted photographs. The idea of automatic plotting from photographs was originated by Major F. V. Thompson, R.E., in 1907, while the pioneer of stereoscopic methods of measuring seems to have been Dr H. G. Fourcade, a South African who read a paper in 1901 to the South African Philosophical Society, describing a stereocomparator of his own invention which enabled the distance of a point to be determined from two photographs taken at the ends of a short measured base.

When the photographs are taken from the ground, the pointings of the camera at the field stations can be measured and recorded. All that is necessary is to carry the camera on a theodolite mounting and fix a telescope to it. These measurements enable the geometrical conditions at exposure to be reproduced in the comparator. When such an apparatus is used with air photographs, however, the directions of the camera axes at the field stations cannot be measured, nor can the position of either field station be fixed independently. Before plotting can begin, the positions of the field stations and the directions of the camera axes must be determined from the photographs themselves. In the first attempts to use air photographs in a plotting machine it was necessary to determine these data by the computation of a resection in space from three or more control points visible on each photograph; an extremely laborious process, and one that did not always yield a very accurate result. Later, methods of adjusting each goniometer* separately to control points by trial were evolved, and until quite recently this was the normal procedure in working with automatic plotters.

There is, however, a serious theoretical objection to any procedure involving independent setting of the goniometers, in that any error in either setting will cause a lack of correspondence in the stereoscopic view, and there is no means of ascertaining which of the two independent settings is in error.

---

* The word "goniometer" is used to denote a projector which is a dimensional replica of the air camera.

In 1926 or thereabouts, Fourcade, whose work on this subject has not received the recognition it deserves, produced a novel design for a stereogoniometer in which the setting of two photographs in exact correspondence could be done without any knowledge of the positions of the air cameras and without the aid of ground points. He showed that if the goniometers were mounted together on a single axis representing the air base, the correspondence setting could be effected by a relatively simple systematic trial procedure; and that when five suitably placed common points had been set in correspondence, the relative positions of the two photographs with respect to one another and the air base had been found. Relating each photograph of a strip in turn to its neighbour and to the intervening air base in this manner, Fourcade worked out a system of aerial triangulation which enabled the relative positions, orientations, etc. of all the photographs in a suitably flown strip to be accurately determined without any reference to the ground at all. The absolute scale and orientation of each could then be determined by adjusting the strip as a whole to a few ground points placed anywhere thereon. His instrument seemed to offer great advantages for automatic plotting as well as for air triangulation, in that the correspondence setting could be made separately and left untouched while the space model was adjusted to ground control. The setting could be left untouched during plotting also, instead of being affected, and perhaps upset, by the plotting movements, as was the case in some of the plotters then on the market. His instrument held out such possibilities for improvement in all kinds of air survey that the Air Survey Committee arranged for two experimental stereogoniometers based on Fourcade's design to be made for the War Office by Messrs Barr and Stroud, and later ordered from the same firm, also for the War Office, an automatic plotting machine designed by Captain E. H. Thompson, R.E.

This plotter, based on Fourcade's principle, is stereoscopic and uses the Porro-Koppé viewing system; that is to say the photographs are carried in holders which are scale replicas of the aerial camera and are viewed from the front through the lens. Bundles of rays from any point in the focal plane emerge from the lens as parallel rays and the point appears in the view as if situated at infinity, as in nature. The apparatus consists of three distinct parts: (1) the photograph carriers, or goniometers, representing the air camera at the two ends of the air base; (2) the stereoscopic viewing system by which the two pictures are combined stereoscopically and examined under magnification; and (3) the plotting mechanism.

The goniometers are mounted on the ends of a rigid bar representing the air base and have five degrees of freedom. Each plate can be rotated in its own plane about the principal point. This is called the "position movement", and represents two degrees of freedom. Each goniometer can be revolved in the plane of the bar about axes perpendicular to this plane and passing through the front node of the camera lens. These are called the "declination movements", and represent a further two degrees of freedom. Finally one goniometer can be rotated relatively to the other round the bar in a plane perpendicular thereto. This is called the

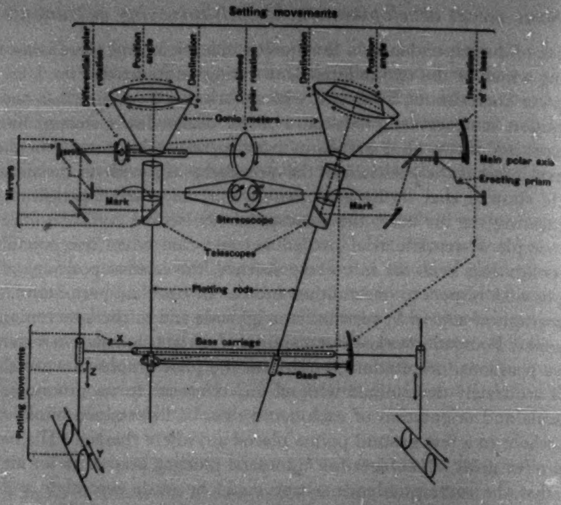


Figure 8. Simplified diagram of the Thompson plotter.

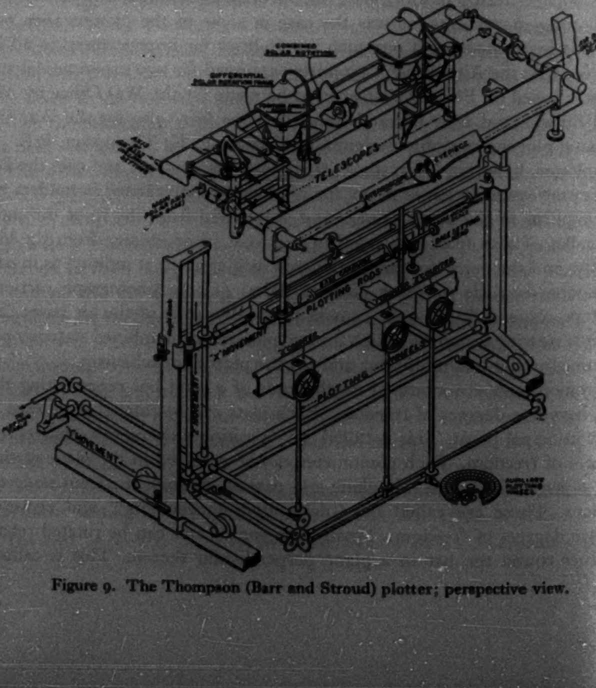


Figure 9. The Thompson (Barr and Stroud) plotter; perspective view.

"differential rotation" and represents the fifth degree of freedom. These five movements are sufficient for the correspondence, or internal, setting. In addition the bar representing the air base can be rotated so that both goniometers revolve together in planes perpendicular to the bar. This movement is required for the external setting, or adjustment to ground control.

The plotting system consists of a base carriage which controls the movement of the plotting pencil and on which the base length is set out to scale; this carriage is movable on rails in the X and Y directions, and is also capable of being raised or lowered in the Z direction. The base carriage is thus movable in the three directions of the co-ordinate axes. To the base carriage are attached by slides the ends of guide rods which represent, or materialize, the rays from any given point seen in the view to its two constituent images on the photographs. At their upper ends the guide rods are attached to the air base by gimbal joints which enable them to pivot in all directions round the front node of the camera lens. The upper part of each rod carries a telescope arranged to look into the camera through the lens, and of which the collimating axis is deflected into the stereoscope by a mirror placed a short distance from the upper joint, the object of this deflection being to bring the two views together into a stereoscopic eye-piece situated conveniently for the operator. The lower part of the rod is always in prolongation of the collimating axis of the telescope. Raising the base carriage (the Z movement) causes convergence of the rods, which are coplanar with the air base in all positions, and makes the floating mark appear to rise. The X and Y movements of the base carriage move the mark from side to side and up and down in the view. These three movements enable the mark to be brought into and kept in coincidence with any point in the view. Suitable gearing transmits the movements of the base carriage to a pencil moving over a plotting table in such a manner that the pencil records the apparent horizontal movements of the mark.

The optical system of the stereoscope comprises three reflecting mirrors and three lenses. It is provided with two types of mark, one similar to that used in the Thompson comparator for correspondence setting, and the other a simple dot to be used for plotting after setting has been completed. It also provides for transposal of the two pictures so as to give a pseudoscopic view, that is to say, to reverse the relief of the picture so that the highest points appear as through they were the lowest and vice versa.

The mirrors M_1, M_2 and M_3 form with F_1, figure 10, a parallelogram in all positions of the plotting rods, and the mechanical arrangements ensure that the mirrors are always perpendicular to the bisectors of the angles of the parallelogram. The nett effect is that in all positions of the rods the visual rays ultimately travel along a fixed path M_3P_1 in prolongation of the polar axis. From P_1 the rays are deflected to the front into the eye-piece—the image being kept upright in all positions of the rods by the AMICI prism P_2, moved by a system of half-speed gearing driven off the polar axis. The marks are engraved on glass plates fixed at F_2 in the telescope. The remaining lenses, prisms, etc. bring the image to a focus in the focal plane of the eye-piece.

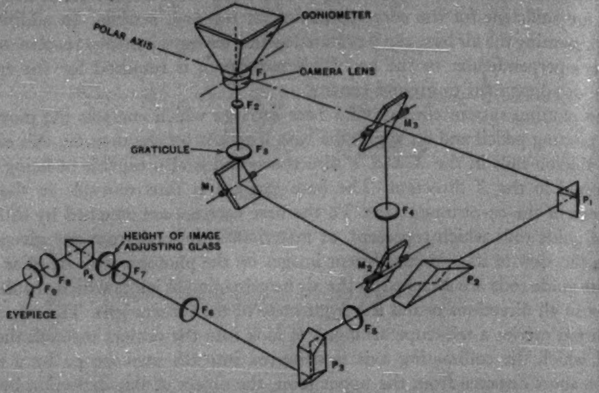


Figure 10. Diagram of optical system for plotter type Z.A. 2.

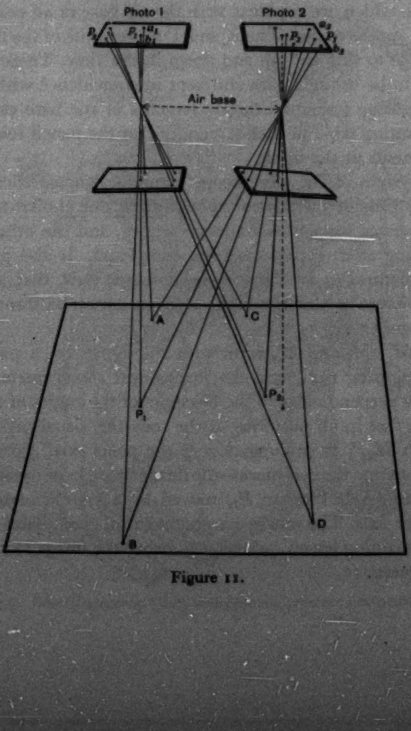


Figure 11.

The internal setting is done by trial, and by successive approximation. Four points at the corners of each overlap (e.g. *A*, *B*, *C* and *D* in figure 11) and the two principal points are examined in turn. Correspondence in the *X* direction can be secured by converging or diverging the plotting levers, but lack of correspondence in the *Y* direction, indicated by splitting of the floating mark, must be corrected by the setting movements. The way in which the floating mark splits up, however, indicates which movements to use. Naturally the correspondence setting requires a certain amount of experience and practice. It may take anything from 5 to 20 min. according to the amount of tilt in the two photographs. The setting movements allow for a divergence of 15° in either goniometer from the central position, while provision is made in the instrument for setting the air base at an inclination of 6° to the horizontal.

In this particular instrument the principal distance of the goniometers is 5 in., as the plotter is designed to work with photographs taken by the F. 24 camera and the Ross E.M.I. lens. If another type of aerial camera had to be used, a different pair of goniometers would have to be fitted.

The external setting may be done in several ways. The first stage is to orient the space model approximately. This is done in the first instance by turning the paper on the plotting table. Next an approximate scale adjustment is made on the base carriage. The air base of the instrument being a fixed length, the scale of plotting is regulated by the device known as the Zeiss parallelogram (because it was first used by Zeiss in their plotter called the "stereo-autograph"), and used in nearly all automatic plotters.*

The scale having been adjusted approximately on the base carriage in the *X* direction, the space model is moved in the *Y* direction by means of the combined rotational movement until the scale error is equal in both directions when it is corrected on the base setting. If no correct position can be found in this way, the

---

* The device may be explained as follows; see figure 12. Suppose *AB* is a base and *P* a point fixed by intersecting rays from its ends. For plotting purposes *B* can be moved to any other point *B′*, provided that the ends of the plotting rays *AP*, *B′P′* (whether represented by rods or by visual rays) are connected and controlled by a distance piece *PP′* which remains in all plotting positions equal and parallel to *BB′*. Since *BPP′B′* remains a parallelogram in all positions, *B′P′* will always be equal and parallel to *BP*. Hence *P′* will follow exactly the movements of *P*. It can be seen then if *A′P* is the base carriage and is made equal to *AB′*, then by similar triangles *A′P′* is equal to the base *AB* and *P′Q′* to *BQ*. The scale of plotting, corresponding to a base length *AB*, can therefore be given by setting off *A′Q′* equal to *AQ* and *Q′P′* equal to *QB*. This principle can be applied in a variety of ways. *B′* for example can be placed on the other side of *A* to *B* or at any other convenient point. *A* also may be moved as well as *B*.

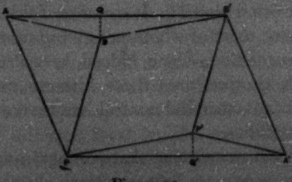


Figure 12.

inclination of the air base must be altered and the necessary corrections must be made either on the base setting or by means of the rotational movement.

Although there has not yet been time to obtain any practical experience of the working of this plotter, it is hoped that it will enable preliminary settings to be done decidedly more quickly than on any continental machines. The latter for the most part have been adapted from plotters designed to work with ground photographs, in which the setting of the goniometers was done from angular and linear measurements made *in situ*; or have been evolved from aerial plotters in which the air positions and the camera pointings had to be calculated independently from resections in space. The objection, from the point of view of stereoscopic correspondence, to independent goniometer settings was probably not fully realized until Fourcade's paper was published in 1926. At all events since then many foreign plotters have been modified, and in some cases entirely redesigned, so as to allow a preliminary correspondence setting to be made before the external setting begins. Most of these instruments, however, still show traces of their ancestry in the retention of what appear to be unnecessary or redundant movements, which do not seem to be of assistance in the setting, but must add substantially to the cost.

All these foreign instruments—and the British instrument too—are primarily designed for detail plotting and contouring, starting from a relatively dense trigonometrical control. A number of surveys, on scales as large as 1/1000, have been made with them, and it is said that they have proved not only more accurate but more expeditious than surveys made on the ground.

I am disposed to think, nevertheless, that most existing plotters will eventually be considerably modified, at any rate for systematic surveying—that is for the mapping of a whole country, as opposed to the mapping of small detached properties. When the photographic conditions can be controlled, a universal instrument will probably prove to be extravagant, and it is likely to be cheaper to eliminate tilt, as far as possible, when the photographs are being taken, or alternatively to photograph at fixed tilts, than to use a plotter which is designed to allow for and correct for random tilts of almost any magnitude.

Again, it is a well-recognized principle of surveying that the fixing of control should be separated from the plotting of detail. For control purposes it is only points that have to be fixed, and the accuracy of their fixing must be substantially greater than that required for detail plotting. Hence a machine that is accurate enough for control work is almost sure to be needlessly accurate, and therefore needlessly expensive, for detail plotting. It seems as if this is now being realized by foreign instrument-makers, most of whom are giving more and more attention to the problem of aerial triangulation. Some, indeed, have already produced simplified instruments, very much cheaper than those hitherto constructed, to do detail plotting after the basic trigonometrical control has been sufficiently elaborated by the more complicated machines.

In conclusion it is worth mentioning that photogrammetry, as it is called, and photogrammetrical instruments, are being increasingly used for a variety of purposes besides land surveying. It can hardly be doubted that they have a considerable future, not only inside but also outside the surveying field.

§ 13. FOREIGN APPARATUS

It may be of interest to give a brief description of two or three of the most important foreign instruments. The easiest to understand is the Gallus-Ferber apparatus, already mentioned, which works on the blink or coincidence principle. The correspondence setting is particularly easy on this machine, since lack of correspondence is shown at once by the direction of the apparent shaking of the picture.

The Zeiss machine, known as the "stereo-planigraph", is stereoscopic, and like the Barr and Stroud instrument uses the Porro-Koppé principle. The marks are inscribed on mirrors on the base carriage, which coact with an automatically focusing lens system to send the rays in the X direction into the stereoscope, but since the marks share in the movements of the base carriage, and the observer cannot follow these movements with his body, the visual system includes an optical shears or scissors between the measuring marks and the eye-piece, a somewhat complicated optical arrangement which must add considerably to the cost of the instrument. The two goniometer lenses and the two marks form the corners of a Zeiss parallelogram. The base setting, applied to the goniometers in earlier models, is done, on the latest models, on the base carriage as in the British instrument. The X and Y movements are given to the base carriage and the Z (vertical) movement to the goniometers, which are carried on a bar parallel to the X movement, and are separately rotatable round the bar and in the declination direction. In the latest (1930) model the two goniometers can in addition be rotated together round the bar also. The earlier models were intended for oblique as well as vertical photography, and they provided for an interchange of the X and Y co-ordinate axes; the latest model is intended for nearly vertical air photographs only, as it is considered that ground photographs and high obliques can be sufficiently dealt with by the stereo-autograph, a much cheaper instrument.

The earlier Wild (Swiss) Autograph also made use of the Porro-Koppé principle, but in the latest model the Porro-Koppé principle has been abandoned and the visual rays are represented, throughout their length, by extensible metal rods pivoting on the perspective centres and each carrying at its upper extremity a prism which reflects the image of a small section of a photograph, placed just above it, to one side into the viewing system. The rays thus reflected are ultimately brought to a focus on two marks located in the focal plane of the stereoscope eyepiece. The photographs are carried in holders which reproduce the dimensions and geometry of the aerial camera, but the photographs are not viewed through the camera lens. Several advantages are claimed for this arrangement, among others that cameras of different focal lengths can be used without the holders having to be changed. The holders or goniometers are carried on a bar representing the air base as in the Thompson and Zeiss instruments, and are rotatable independently in two directions as well as together. The bar also is rotatable in its own plane in two perpendicular directions: this is presumably to allow for interchange of axes when horizontal photographs and high obliques are being plotted. The plotting rods are attached to a base carriage in which provision is made for base settings in the *Y* and *Z*, as well as in the *X* directions. The *X* and *Y* movements are given to the base carriage, and the *Z* movement to the goniometers as in the Zeiss instrument.

There are a number of other instruments on the market besides the above, the Santoni (Italian) and the Poivillers (French) instruments being the most important.

CN AND AlO BANDS IN THE STUDY OF THE CARBON ARC

By N. R. TAWDE and S. A. TRIVEDI
Physics Laboratory, Royal Institute of Science, Bombay

Received 5 *December* 1938. *Communicated by W. Jevons. Read in title* 28 *April* 1939

ABSTRACT. A systematic study of the problem of the simultaneous excitation of two band systems in the same source has been made by measuring the gross intensity-distribution in the AlO blue-green and CN violet systems in the carbon arc. The results have been applied to the interpretation of the relation between the vibration temperature and rotation temperature in this particular source, in the light of the results obtained by Ornstein and Brinkman. The ratios of the excitation temperatures of the CN system to those of the AlO system measured simultaneously in the same source have been examined and compared with corresponding ratios for C_2 to CH derived by Lochte-Holtgreven under similar conditions. It has been concluded that in the case of simultaneous excitation, so long as the molecules remain the same, the ratio of their temperatures remains practically constant, irrespective of the conditions.

§ 1. INTRODUCTION

The investigation of the simultaneous excitation of spectra of two different molecules in the same source and its bearing on the nature of excitation and the distribution of molecular populations and temperatures within the source is of comparatively recent growth, and has not yet led to definite conclusions. For instance, Lochte-Holtgreven's(1) measurements on C_2 and CH bands in a discharge through acetylene gave excitation temperatures differing by over 2000° K. for the same discharge tube under uniform conditions. When, however, ethylene was used instead of acetylene in the discharge, different temperatures were obtained but with a difference of about the same order(2). On the other hand, van Wijk(3) has found a common effective temperature for N_2^+ and N_2 (second positive) bands excited in the same discharge. All these results are derived from the energy-distribution among rotational levels of one or more bands. As different investigators have reached divergent conclusions, we have thought it desirable to undertake similar investigations from the standpoint of vibrational energy. The AlO and CN molecules have been selected for this purpose, as they give clearly defined band systems whose constants are known with certainty. When excited in the carbon arc, the blue-green system ($^2\Sigma \rightarrow {}^2\Sigma$) of AlO and the violet system ($\mathrm{B}^2\Sigma \rightarrow \mathrm{X}^2\Sigma$) of CN can be observed simultaneously. Moreover, no accurate quantitative data on the gross intensity-distribution in the AlO system is so far available, though we know enough of the CN system. The carbon arc was selected as the source in view of the large amount of available data which would enable us to put our results to the test, particularly for the comparison of vibration temperature with rotation temperature in band spectra. We have published in advance a brief note$^{(4)}$ on the latter subject. It is proposed to give here the details of the investigation and to discuss some points that appeared to be interesting.

§ 2. EXPERIMENTAL

The problem of the study of energy-distribution in a band system reduces in practice to the quantitative estimation of band-intensities. The method adopted and the details of the technique of measurement have been dealt with and discussed at length by Johnson and Tawde$^{(5)}$ and Tawde and collaborators$^{(6),(7)}$. The bands have been photographed on a Hilger constant-deviation spectrograph and a Fuess quartz prism instrument. The sequences $v'-v''=-1$ and -2 of the CN system were recorded on the Hilger instrument along with the AlO bands, and when these were again photographed on the quartz instrument it became possible to express both systems on the same scale of intensities. The observed intensities were expressed in terms of the energy-distribution of a standard lamp which was previously calibrated. The microphotometer used was a self-recording Zeiss instrument of high precision. In spite of all the precautions taken in the photometric methods, the probable error in the final values may extend to the order of about 10 per cent which is not very high in work of this type.

§ 3. RESULTS

In tables 1 and 2 are given the values found for the AlO and CN bands respectively; they are expressed on the basis of 100 units of intensity for the (0, 0) band in each case. Those for the CN bands are converted to these values after they have been obtained on the basis of AlO results.

In table 3 are recorded the values of $\Sigma\, I/\nu^4$ for upper vibrational levels in each of the band systems.

From the values of statistical weights of each of the upper vibrational levels, which are proportional to the numbers $\Sigma\, I/\nu^4$ given in table 3, the excitation temperatures have been derived as follows:

For the AlO system: 3450° K.
For the CN system: 6200° K.

§ 4. CONCLUSIONS AND DISCUSSION OF RESULTS

The intensity-distribution in the CN system has been exhaustively studied and discussed by previous authors$^{(8),(13)}$ and our present conclusions generally confirm their observations. However, the investigation enables us to compare the CN system with the AlO system. From theoretical considerations, in a band system, the shape of the Condon parabola (i.e. the separation between its two branches) is known to be a function of the difference $(r_e'-r_e'')$ in the internuclear distances in the upper and lower electronic states; this accounts for profuseness or otherwise of bands in a system. In the CN and AlO systems under investigation, the values of $(r_e'-r_e'')$ are $-0{\cdot}021$ A. and $+0{\cdot}049$ A. respectively. Therefore, the CN system should give a much narrower parabola and smaller number of bands than the AlO system. This is true from the quantitative measurements that we have made and interpreted on the basis of the Franck-Condon theory of most probable transitions. The CN and AlO systems contain respectively 21 and 34 observable bands, the change of sign of $(r_e'-r_e'')$ being associated with the reversal of the branch $r_{\min}\rightarrow r_{\min}$ in CN to $r_{\max}\rightarrow r_{\max}$ in AlO and vice versa. In both the systems, the experimental results have given a definite proof that potential energy of nuclear vibrations in the upper and lower electronic states is governed in general by expressions due to Morse⁽⁹⁾ and Rydberg⁽¹⁰⁾. This is unlike the conclusion reached from observations on N_2 (second positive), BeO and C_2 (Swan) by Tawde⁽¹¹⁾, Johnson and Dunstan⁽¹²⁾ and Tawde and collaborators⁽⁶⁾,⁽⁷⁾. In these molecules, Rydberg's expression is certainly the better of the two, the derivations from Morse's expres-

Table 1. The AlO system

Bands	Intensity I	I/ν^4	Transition probability
3, 0	17·9	6·19	0·150
4, 1	7·22	2·55	0·096
5, 2	5·29	1·90	0·120
6, 3	3·66	1·34	0·111
7, 4	3·13	1·17	0·357
2, 0	22·5	9·00	0·137
3, 1	12·2	4·97	0·120
4, 2	9·14	3·85	0·145
5, 3	8·23	3·50	0·221
6, 4	7·15	3·09	0·258
7, 5	4·77	2·10	0·642
8, 6	3·01	1·34	—
9, 7	1·54	0·69	—
1, 0	40·5	19·0	0·240
2, 1	26·8	12·8	0·194
3, 2	17·9	8·72	0·211
4, 3	15·0	7·44	0·280
5, 4	8·47	4·27	0·270
6, 5	5·71	2·93	0·244
0, 0	100	55·1	0·478
1, 1	29·6	16·6	0·210
2, 2	20·1	11·5	0·175
0, 1	45·0	30·0	0·260
1, 2	35·8	24·3	0·308
2, 3	27·9	19·3	0·293
3, 4	18·2	12·8	0·309
4, 5	9·93	7·07	0·266
5, 6	8·48	6·11	0·387
6, 7	6·32	4·61	0·385
0, 2	36·8	30·0	0·260
1, 3	23·0	19·0	0·240
2, 4	15·6	13·1	0·198
3, 5	10·1	8·60	0·208
4, 6	6·50	5·59	0·210

sion being widely divergent in one of the branches, viz. $r_{\min} \rightarrow r_{\min}$. It has to be seen whether the similarity and agreement, shown in these respects by CN and AlO systems, have any connexion with the similarity of their electronic states and their multiplicity or with simultaneous excitation within the same source. Evidence so far available shows that the conditions of excitation make no difference in the transition probabilities. These points require theoretical examination from the standpoint of the existing law of force, the limits of its validity, and the consequent derivation of an accurate potential energy expression in closer agreement with Rydberg's empirical form. In the study of these problems, it has to be remembered that disagreement between theory and experiment exists, if at all, for relatively large positive values of $r - r_e$, i.e. for the higher vibrational states.

Table 2. The CN system

Bands	Intensity I	I/ν^4	Transition probability
1, 0	23·6	3·88	0·181
2, 1	30·6	5·01	0·325
3, 2	2·55	0·41	0·062
0, 0	100	22·6	0·759
1, 1	50·8	11·3	0·529
2, 2	25·2	5·56	0·360
3, 3	13·2	2·90	0·440
4, 4	11·4	2·47	0·510
0, 1	14·0	4·40	0·148
1, 2	12·3	3·78	0·176
2, 3	8·95	2·71	0·175
3, 4	5·06	1·51	0·229
4, 5	4·44	1·31	0·270
5, 6	2·65	0·78	0·453
0, 2	6·12	2·74	0·092
1, 3	5·47	2·39	0·111
2, 4	4·98	2·13	0·138
3, 5	4·20	1·76	0·267
4, 6	2·59	1·06	0·219
5, 7	2·33	0·94	0·546
6, 8	0·99	0·40	—

Table 3. Excitation functions $\Sigma I/\nu^4$

AlO system		CN system	
Progression v'	I/ν^4	Progression v'	I/ν^4
0	115·14	0	29·71
1	78·94	1	21·38
2	65·69	2	15·41
3	41·26	3	6·58
4	26·50	4	4·84
5	15·78	5	1·72
6	11·97	6	0·40
7	3·27		
8	1·34		
9	0·69		

§ 5. STUDY OF THE CARBON ARC

It is mainly in respect of the temperatures in the arc that the present study of CN and AlO bands has yielded some fruitful results. These will be discussed under three heads: (*a*) the examination of temperature equilibrium within the arc; (*b*) the comparison of vibration temperature with the rotation temperature; and (*c*) the temperature-distribution within the arc.

Under the first head, we have first to examine whether the fall of intensity obtained along sequences and along the different v'' progressions is as expected if temperature equilibrium in the initial state is assumed. For the condition to hold, log (I/ν^4) should be related to vibrational energy in the initial states by a straight-line graph. These relations are plotted in figure 1 for progressions, and in figure 2 for all the sequences of AlO system. From the nature of the graphs, we can safely conclude that emission of AlO bands is a process which is dependent upon statistical equilibrium of molecules in the initial state at one definite temperature; the temperature evaluated from figure 1 is 3450° K. Similar calculations for CN bands give the value of 6200° K. for the temperature. We have next to examine whether these excitation temperatures correspond to true temperatures within the arc, in order to show the existence of temperature equilibrium of excited molecules with

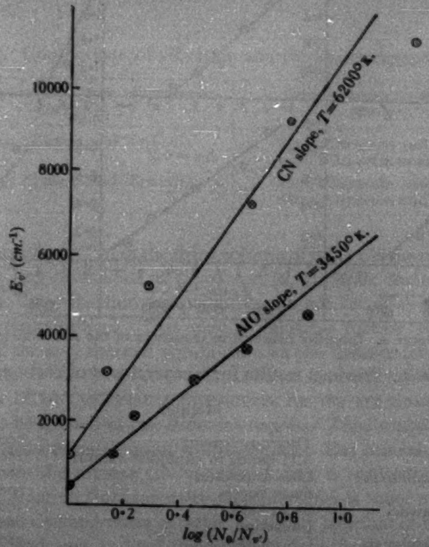


Figure 1

the surroundings. Table 4 shows the results for temperatures of carbon arcs derived by different methods by previous authors. From this table we can conclude that the temperatures derived by us are within the range observed by many other investigators, i.e. in the neighbourhood of 6000° K. for the brightest portion. We have reasons to believe that the temperature is about the same at all distances from

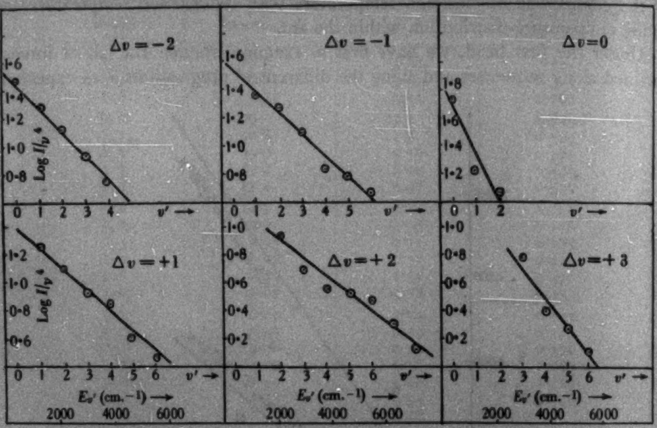


Figure 2. Intensity relations in sequences of the AlO system.

Table 4. Previous results for temperatures of carbon arcs

Author	Method	Temperature
Mathiesen	Thermo-element	3000° K. } Considered
Nagashima	AlH bands in exploded vapour	2500° K. } too low
Ornstein and Sambursky	Line intensities	6000° K.
Johnson	Swan bands	6000° K.
Johnson and Tawde	Swan bands	5000° K.
Mankopff	Ionization	6000° to 8000° K.
Ornstein and Brinkman	CN bands	5870° K.
Krygsman	Swan bands	6000° to 6500° K.
Tawde	CN bands	7000° K.
Hörman	Electron density	6000° to 7000° K. along axis of arc

the electrodes along the arc axis. While the observations of Ornstein and Brinkman(14) support this view, it is difficult to accept their conclusion that it does not change with current. It may be that such a change, if any, is either too small to be detectable by the methods used or else that the changes along axial length occur too abruptly to maintain uniform conditions during the time of exposure. It is not improbable, however, that a much higher temperature, amounting to 7800° K., does exist along the axial line as observed through holes in the electrodes by the above-mentioned authors. There is Hörman's(16) estimate of temperature along the axis, namely 6000° to 7000° K. Immediately around the axis, if temperatures lower than these are assumed, a true average of 6000° K. observed *across* the arc axis is highly probable under equilibrium conditions. The agreement, within experimental error, of our CN temperature with the results obtained by several independent observers makes it likely that this is the true temperature of the arc. This we take as evidence of temperature equilibrium of radiating entities with the surroundings.

In discussing the vibration and rotation temperatures we are guided by the results obtained by Ornstein and Brinkman(14) and Ornstein and van Wijk(15) from the distribution of rotational energy in the arc. Their values are compared with ours in table 5.

Table 5. Comparison of vibration and rotation temperatures

Bands	Temperatures from vibrational energy (° K.)	Temperatures from rotational energy (° K.)
CN	6200 (Tawde and Trivedi)	6500 (Ornstein and Brinkman) 5000 (Ornstein and van Wijk)
AlO	3450 (Tawde and Trivedi)	3275 (Ornstein and Brinkman) 2800 (Ornstein and van Wijk)

The temperature of 3275° K. in the third column of table 5 is the mean of 3600° K. and 2950° K. from line intensities of (1, 0) and (0, 0) bands respectively of AlO. Comparison shows that vibration temperatures are not far away from the corresponding rotation temperatures. This may be taken as further evidence that, at least in the case of the arc, thermal equilibrium at one definite temperature exists for vibrational states as much as for rotational states, and simultaneously.

The derivation of two different temperatures for the arc shows that the two band systems must be excited in two different regions. This brings us to the study of the radial temperature-distribution within the arc. The central gaseous zone is the place from which the intense CN violet radiation is emitted; its temperature according to our estimation from intensity-measurements is 6200° K. The fact that the AlO system gives a temperature of 3250° K., nearly half that of the central part, means that these bands are excited almost within the flame part, well removed from the arc axis. There is evidence for the existence of the latter zone from the appearance of the blue-green colour of the flame and from the observation that the bands are suppressed if only the central violet glow is focused on the slit of the spectrograph. It may therefore be said of arc excitation in general, that it causes the emission of one system or another, only when the respective molecules find themselves in the appropriate zone in which temperature-equilibrium is attained, and the intensity-distribution then belongs to that temperature. If we could identify and locate the positions of different molecular spectra of varying excitation temperatures by their characteristic visible colours in the arc, it would offer us a means of determining the radial temperature-slope by the application of band-spectroscopic methods.

The above reasoning cannot be extended to explain the results of Lochte-Holtgreven[(1, 2)] on the simultaneous excitation of C_2 and CH bands in the same discharge. The excitation temperatures of 4700° and 2000° K. respectively for C_2 and CH bands with acetylene in the discharge, and 3800° and 1600° K. with ethylene, show that they are too high for the source used. This would give, in the first place, absence of statistical equilibrium with the surroundings, and the discharge cannot be separated into well-defined temperature zones as in the case of arc. Such large divergence in the temperature values from acetylene and ethylene is probably due to complicated dissociative processes resulting from chemical action in the discharge. One interesting point of similarity between these observations and our conclusions is that the ratio between excitation temperatures of the respective molecules under different conditions is nearly the same. Thus:

Ratio of C_2/CH temperatures:

2·35 with acetylene in discharge (Lochte-Holtgreven).
2·37 with ethylene in discharge (Lochte-Holtgreven).

Ratio of CN/AlO temperatures:

1·90 from rotational energy (Ornstein and Brinkman).
1·78 from rotational energy (Ornstein and van Wijk).
1·8 from vibrational energy (present paper).

This naturally leads us to the conclusion that so long as we resort to the same pair of molecules, any change in temperature in one molecule of the pair, as we pass from one condition to the other, causes a proportionate change in the other.

§ 6. ACKNOWLEDGEMENTS

We are much indebted to the authorities of the Science College, Patna, for allowing us the use of their microphotometer and extending their kind hospitality to one of us (S. A. T.) during our stay at Patna.

REFERENCES

(1) Lochte-Holtgreven. *Z. Phys.* **64**, 443 (1930).
(2) Lochte-Holtgreven. *Z. Phys.* **67**, 590 (1931).
(3) van Wijk. *Z. Phys.* **59**, 313 (1930).
(4) Tawde and Trivedi. *Nature, Lond.*, **140**, 463 (1937).
(5) Johnson and Tawde. *Proc. Roy. Soc.* A, **137**, 575 (1932).
(6) Tawde and Desai. *Proc. Ind. Acad. Sci.* **6**, 266 (1937).
(7) Tawde and Patel. *J. Bombay Univ.* **6** (ii), 29 (1937).
(8) Tawde. *Proc. Ind. Acad. Sci.* **3**, 140 (1936).
(9) Morse. *Phys. Rev.* **34**, 57 (1929).
(10) Rydberg. *Z. Phys.* **73**, 376 (1931).
(11) Tawde. *Proc. Phys. Soc.* **46**, 324 (1934).
(12) Johnson and Dunstan. *Phil. Mag.* **16**, 472 (1933).
(13) Ornstein and Brinkman. *Proc. K. Acad. Amst.* **34**, 33 (1931).
(14) Ornstein and Brinkman. *Proc. K. Acad. Amst.* **34**, 498 (1931).
(15) Ornstein and van Wijk. *Proc. K. Acad. Amst.* **34**, 44 (1931).
(16) Hörman. *Z. Phys.* **97**, 539 (1935).

REVIEWS OF BOOKS

The Mobility of Positive Ions in Gases, by Prof. A. M. TYNDALL. Pp. ix + 93. (Cambridge Physical Tracts. Cambridge University Press.) 6s. net.

As it is becoming increasingly difficult for one person to read all the original papers published in more than a few branches of physics, it is important that authoritative surveys of special subjects should be given from time to time in a form suitable for the non-specialist. In the volume under review Prof. Tyndall has done this in an admirable way for the work on positive ion mobilities carried out mainly by the Bristol School of research workers. By avoiding too much detail he has given a very clear account of his subject which will be valuable to those whose main interests lie in other directions. All those who are contemplating the use of metallic ion sources will find his experiences of these given in chapter III very helpful. The book can be strongly recommended as giving a clear and concise account of the subject with which it deals.

E. T. S. W.

Der Smekal-Raman Effekt. Ergänzungsband 1931–1937, by K. W. F. KOHLRAUSCH. Pp. x + 288. (Struktur und Eigenschaften der Materie, XIX.) (Berlin: Julius Springer, 1938.) Price RM. 24.

In this supplementary volume the same order of treatment is followed as in the original volume. Most sequels are disappointing; this book is an exception. It has an excellent account of the theory of molecular vibrations, and contains tables for finding the numbers of vibrations in each fundamental normal mode for the different symmetry groups. These tables have been synthesized from Placzek's tables in the *Handbuch der Radiologie*, and Brester's monograph on the infra-red. This has led the author into a slight slip in the tables for D_{3h}, D_3, and C_{3v}, since he lists two different subsidiary digonal axes in the first two cases, and two different sets of vertical planes in the last. Actually this is not likely to lead to any numerical difficulty. The combined tables so formed are exceedingly useful to infra-red and Raman spectroscopists who wish to deal with the vibrations of molecular models and the Raman effect in liquids. For crystals they are not so useful, but no tables especially suitable for crystals have yet appeared.

The formulae for the frequencies of vibration of the different molecular models in the original volume were for the greater part in terms of central forces. In the new volume, a section is devoted to listing the frequency formulae for valence-force models. Much is compressed here into a small space, but one could wish that a little more space had been taken to make clear the accompanying diagrams. The discussion of the properties of the Rayleigh lines, Raman lines, and influence of intermolecular forces is excellent. There is much space devoted to the assembling of the results for comparing the predicted behaviour of molecular models with the actual Raman spectra observed, a study which the author and his school have done so much to further.

The attempt to list the results obtained for various substances has wisely been abandoned, but the new references (nearly 900!) have been cross-indexed under substances. The index is very well done. There are slips in the book, but it is amazing that there are not a good many more; so much information is packed into it. No worker in this field should be without it.

A. C. M.

Electron Optics, by OTTO KLEMPERER. Pp. x + 107. (Cambridge University Press.) Price 6*s*. 0*d*.

Electron Optics, by L. M. MYERS. Pp. xviii + 618. (London: Chapman and Hall, Ltd.) Price 42*s*. 0*d*.

More than four years have elapsed since the publication of Brüche and Scherzer's *Geometrische Elektronenoptik*, and in view of the present technical importance of cathode-ray tubes, and the possibilities of electron-optical systems like the electron microscope, the appearance of an adequate textbook in English may be said to have become overdue. The volumes now under review are both addressed to those interested in the practical applications of the subject, rather than to theoretical physicists, and neither of the books assumes any previous detailed knowledge of optics on the part of the reader.

Dr Klemperer's account is a model of brevity, and is content to quote formulae where necessary; he modestly entitles it to a compilation, and naturally gives emphasis to the practical and experimental methods, often of an original character, by means of which the design of electron-optical lenses has been studied in the laboratories of Messrs Electrical and Musical Industries, Ltd. In addition there is a useful and concise account of the aberrations of an electron-optical system. The principle of least action (attributed here to Lagrange) is mentioned, but no attention is given to Hamiltonian methods in the presentation. The book may be commended as a useful contribution to the study of the subject, although some of the methods described are admittedly in the experimental stage and will best be taken as useful suggestions rather than as fully accredited by experience.

In spite of the more or less informal style of the book, it is a pity that the elementary derivation of the formula for the focal length of a magnetic lens is not given a more exact character. The formula on page 54,

$$\ddot{r}=\frac{e^2r}{4m^2}H_z^2,$$

would suggest that the path of the electron is always convex to the axis. Further, the centrifugal force is not balanced by the inward-acting force, as though the particle were attached to a string. Bouwers's paper in *Physica* (1937), which is quoted in the present connexion, was not the first to give an elementary derivation of the lens formula; see, for example, the *Journal of the Television Society* in this country in 1934. The matter may be a rather trivial one, but it is worth something to have it reasonably correct in a formal sense. After all, this is not a recondite problem in cosmogony or wave mechanics.

Mr Myers has aimed at providing a more complete work of reference for the physicist and engineer, and he has not shrunk from including in the volume those parts of the general theory of electric and magnetic fields which are necessary for the derivation of the working equations. Apart from this, the book covers much the same kind of ground as Brüche and Scherzer's pioneer work, although the principal advances since 1934 are fully dealt with, and the treatment is of an even more expository character. Here again, after a short discussion of the wave surface and the eiconal, the author prefers to restrict himself to the dynamical method, and to outline the treatment of the aberrations due to Scherzer, rather than to follow Glaser in using the mixed eiconal. If only these fundamental optical conceptions had received more attention in the past by persons writing for the sake of instructing, they would be more readily available for use in such an instance as this, in which, when once the necessary and sufficient variables for the expression of the eiconal have been found, the general character at least of the electron-optical aberrations can be reached in a surprisingly simple manner. However, the calculation of the aberrations is hardly a matter which will concern very many persons.

The book is written in a clear style and appears to be generally accurate, although the present writer has yet not been able to study more than some parts of its six hundred odd pages. There are a few misprints here and there in formulae; they seem almost inevitable in a book of this size. In the necessarily extensive discussion a few points have been misunderstood; for example, surely the spreading of an originally parallel beam of electrons is simply due to mutual electrostatic repulsions? Wires carrying parallel currents attract each other, but, as far as magnetic fields are developed by fast free electrons, it has been explained on the basis of the theory of relativity that the mutual forces could in effect be measured by an observer moving with a pair of electrons, and as the magnetic fields would not be apparent in this observer's system there could be no attraction.

There are very full and useful descriptions and illustrations of many of the modern electron-optical devices such as electron multipliers, television tubes, and so on to the development of which the author has himself contributed as a member of the resourceful research staff of the Marconi Company. A brief section at the end deals with the application of electron-optical principles to radio tubes. There is also a chapter on vacuum technique. A fairly full section deals with the electron microscope, and summarizes very fairly the present state of practice in this connexion.

It will thus be understood that the practical sections are the more original parts of the work, and the book will undoubtedly be very valuable to any one who wishes to obtain a survey of current practice, without concerning himself over-much with theoretical matters. Mr Myers will deserve the gratitude of many for bringing together so much information which has not hitherto been available in book form. No doubt the technique of electron optics is in a state of rapid development, but many of the main features of the subject are sufficiently well established to justify the appearance of this useful volume.

L. C. M.

Experimental Methods in Gas Reactions, by A. FARKAS and H. W. MELVILLE. Pp. ix + 389. (Macmillan and Co.) 30*s.* net.

The authors of this book state in the beginning of their preface: "Within recent years the study of the kinetics of gas reactions has assumed a prominent role in chemistry. One of the consequences of that development has been the appearance of a large number of monographs dealing with every aspect of the subject. Nearly all the texts, as is natural, are confined to a discussion of the results of experiments rather than to the methods whereby the data have been obtained. The present book is an attempt to remedy the matter and repair an obvious omission". This is an understatement both of the need and the authors' attempt to fill it. Far more workers than those who are primarily engaged in the study of gas reactions have felt the need of a book dealing with the experimental technique of investigations on gases both in the laboratory and in industrial application and have had great difficulty in acquiring the necessary information. The attempt of the authors to fill the gap has been more than successful and the reviewer has been astonished to find how much information can be clearly presented in less than 400 pages.

The first chapter deals with the physical theory of gases and with the laws of chemical kinetics. All the leading relationships are succinctly stated and tables of constants relating to a large number of gases are included. The presentation is in a very convenient form for reference when an appeal to theory is necessary to solve a practical difficulty. In chapter II, apparatus and methods for the control and measurement of temperatures and pressures are dealt with from near absolute zero to high furnace temperatures, and from high vacua to pressures of hundreds of atmospheres. That this section should be incomplete is inevitable within the scope of a book of this size, and there is no doubt that specialists in one branch or another will find gaps, but to the general practitioner it is a mine of information. The third chapter deals with the preparation and analysis of gases and is very comprehensive, excluding however ordinary macroanalysis which, as the authors say in their preface, has already been adequately dealt with in several excellent books.

Chapter IV deals with photochemical technique and will be of great value to anyone taking up this rather specialized branch. The last chapter deals with the experimental methods for the investigation of chemical reactions with descriptions of the apparatus required and is very wide in its scope.

Throughout the text full references to original papers are given in a most convenient form on each point discussed, and in the appendix there is a bibliography of the leading works on each subject. The copious diagrams of apparatus, placed conveniently in the text referring to them, are very clearly reproduced and not overburdened with non-essential details. There are good subject and author indexes, and the printing and production of the book is up to the usual high standard one expects from the publishers. The labour of the authors in sifting and selecting the most useful material from the vast and scattered literature must have been immense, but readers may accept the result with confidence when the work has been done by two such well-known experimenters as Dr Farkas and Dr Melville. The book will be a great labour-saver to all those who are doing any experimental work on gases.

A. B.

Laboratory Experiments in Elementary Physics, by N. H. BLACK. Pp. xiii + 264. (New York: The Macmillan Co., 1938.)

This book gives directions for the performance of some 60 experiments in a first-year course of practical physics. Instructions are divided in such a way that the first part refers to the more fundamental part of an experiment—which could be completed in about 50 min. The second part is more or less optional, depending on time available, equipment, etc.

In the introduction there are some notes on accuracy of measurements and their relative importance.

A good illustration is found when the ordinary general specific heat equation is written in the form

$$m\,(t_m - t_{\text{mix}})\,X_1 = \omega X_2\,(t_{\text{mix}} - t) + m_c X_c\,(t_{\text{mix}} - t),$$

and each term is worked out separately. This is much better than the form given by the author (the liquid being water),

$$m\,(t_m - t_{\text{mix}})\,X = (\omega + 0{\cdot}1c)\,(t_{\text{mix}} - t_\omega),$$

in which the units are not consistent. Can this be because we are twice told that specific heat is found to be "— calories"? No! It is because the term "thermal capacity" has not been used.

J. H. B.

Modern Developments in Fluid Dynamics. (Oxford Engineering Science Series.) An account of theory and experiment relating to boundary layers, turbulent motion and wakes. Composed by the Fluid Motion Panel of the Aeronautical Research Committee and others and edited by S. GOLDSTEIN. 2 vols: vol. 1, pp. xxii + 330; vol. 2, pp. xi + 331–702. (Oxford University Press.) 50s.

The present work deals with some novel and difficult branches of the theory of the motion of actual fluids of small viscosity; in particular with the laminar boundary layer and with turbulent flow in boundary layers and pipes. In contrast to classical hydrodynamics which bore comparatively little relation to the actual motion of fluids, the subject of the present work is illustrated and checked by experiment at every stage.

For the reader who intends to tackle the whole work, the introductory chapters are perhaps unnecessarily long and involve some repetition, though they are well written and admirably illustrated by photographs of actual fluid flow. Some faults of arrangement may be due to the multiplicity of authorship, though the effects of this are minimized by unusually good editing. Thus, chapters IX and XI on the drag of two-dimensional and three-dimensional bodies would be much better subdivided into streamline and bluff bodies respectively. In particular the sphere and cylinder would be better treated together.

The photographs of fluid flow illustrate the much greater difficulties encountered with air than with water, but some of the most recent air photographs taken with the electrical-hot-spot technique go far to rival those taken in water for pictorial effect, while they are superior for such purposes as the measurements of turbulent components of velocity. Together with the use of the hot-wire technique to measure fluctuating velocities in air, and of ultramicroscopic particles in water, they represent a remarkable triumph of the experimenter in this most difficult subject of turbulence. Considerable space is given to experiments on turbulent flow in pipes and channels which are of the earliest date and perhaps the easiest technique; it is remarkable how much of this information has been found applicable to the turbulent boundary layer which is of more direct importance in aeronautical applications.

On the theoretical side the subject of turbulence is very difficult to the beginner, and it is considered that even more space might well have been given to the fundamental parts of the subject, such as the momentum and vorticity transfer theories. Burgers's statistical theory, which one hopes may ultimately supersede all others, is merely mentioned without any formulae. The whole subject is in that most intriguing stage of rapid development in which it is difficult to give an adequate contemporary account. The policy of the book appears to be rather to include references to all original papers of importance than to give a satisfactory self-contained account of the subject.

In contrast to the experimental chapters, some parts of the theoretical developments appear to suffer from insufficiency of explanatory diagrams. Few misprints have been observed, but a reference to frequencies above 100 ft. per second seems to have escaped the editor's notice.

The book should be absolutely indispensable both to the beginner and to the expert, and it is perhaps sufficient praise to describe it as worthy of Horace Lamb to whose memory it is dedicated.

C. N. H. L.

Aircraft Design, by C. H. LATIMER NEEDHAM, M.Sc. (Eng.), Lond., B.Sc., F.R.Ae.S., M.I.Ae.E. Vol. 1, Aerodynamics. Pp. xii+215. (Chapman and Hall, Ltd.) 13*s*. 6*d*. net.

The problems which arise in aircraft design cover a wide field, and a comprehensive treatment of every aspect of these problems, based on the results of research and practical experience, would be a formidable task. A writer of a text-book on aircraft design has therefore to decide at the outset, if he wishes to restrict the book to a reasonable size, whether he proposes to give a broad general outline, or whether he intends to concentrate, in greater detail, on some of the more important aspects of the subject. Mr Needham does not fall between these two stools, for it is clear that he has chosen the first alternative.

To ensure that his treatment shall not be continually interrupted with explanations of the principles underlying the various features of design and with descriptions of devices used on aircraft, the book is presented in two volumes; the first, now under review, outlines the principles of flight and stability, whilst the second, to be ready shortly, will deal mainly with the mathematical treatment of design. Volume 1 has 14 chapters, and deals with aerofoil flow, lift and drag, the control system, stability, the forces acting on an aeroplane, the slotted wing, and auxiliary control and variable lift devices. Arrangement of lifting surfaces, types of aircraft, and the manœuvres of aeroplanes are described, and the airscrew, aeroplane performance, parasitic and interference drags are dealt with. Volume 1 gives a sound treatment, in simple language, of the subjects chosen, and it can be thoroughly recommended not only to students particularly interested in aircraft, but also to those who wish to obtain a general knowledge of the aerodynamic principles underlying the behaviour of an aeroplane.

A. FAGE